VOLUME VII NOVEMBER, 1910 No. 4

MOSQUITOES AND MALARIA.

A STUDY OF THE RELATION BETWEEN THE NUMBER OF MOSQUITOES IN A LOCALITY AND THE MALARIA RATE.

BY H. WAITE, M.A., B.Sc.

PROFESSOR ROSS, in his Report on the Prevention of Malaria in Mauritius*, has dealt briefly, from a mathematical point of view, with the influence of mosquitoes on malaria and he suggested in a letter to Professor K. Pearson that a further treatment from the more mathematical side would be of value. Such is the origin of the present investigation and it is to the above-mentioned Report that I am entirely indebted for the numerical data upon which this enquiry is based.

It has been proved conclusively that malaria is caused by the bite of a mosquito which has already bitten a previously infected person; and that the infection is probably conveyed only by the sub-family of mosquitoes called Anophelina. Hence, in order that the disease may spread in a neighbourhood, there must be present a source of infection in the shape of persons suffering from malaria, and a sufficient number of anophelines to act as carriers. Moreover, since recoveries are constantly taking place, it is clear that the increase or decrease in the number of cases will depend largely on the relative frequency of anophelines and of human beings.

Malaria Rate. The ratio of the number of persons affected with malaria to the total population of a district at a given time is called the *Malaria Rate* of the district at that time. In general, the rate is continually changing owing to (*a*) new infections, (*b*) recoveries, (*c*) emigration and immigration, (*d*) the birth and death rates, and (*e*) the extent to which cases are isolated, as well as owing to changes in the mosquito population.

As emigration and immigration vary considerably in different localities, and in the same locality at different times, their influence on the malaria rate cannot be

* *Report on the Prevention of Malaria in Mauritius*, by Ronald Ross, D.P.H., F.R.C.S., D.Sc., LL.D., F.R.S., C.B., published by Waterlow and Sons, London, 1908.

satisfactorily dealt with except in particular cases where the necessary statistics are available; neither would results in general terms be of much practical use.

It is probable that the risk of infection may depend partly on the age and past history of an individual; for example, a person who has completely recovered from one or more attacks may become partially or entirely immune. So far, however, I have not found any statistics which deal with this phase of the question.

The combined effect of the birth and death rates will, as a rule, be small compared with the effects under (*a*) and (*b*); e.g., the birth rate for Mauritius in 1906 was 2·8 per thousand per month and the death rate 3·3; of the deaths, 1·2 per thousand were due to malaria. These figures do not differ appreciably from the average taken over a long period of years.

I propose, therefore, in the present paper, to confine my attention to the effects on the malaria rate of the first two of the above causes and to leave for a later consideration the minor contributory causes.

I shall also assume in this preliminary survey that both the human population and the anophelines are uniformly distributed over the area in question; that persons of all ages, whether they have previously suffered from malaria or not, are equally liable to infection; that the number of anophelines per unit of the population remains approximately constant during the period under review; and further that infected and non-infected persons are equally exposed to the risk of being bitten.

It is obvious that the results so obtained will require modification in practical application and ought to be taken into account when the theory is developed to a higher approximation; they will be, at the best, tentative and suggestive, but it is hoped that they will tend to throw light on the problems which arise when the specialist has to consider remedies for the disease.

Infection Rate. The number of new infections in a locality during any particular month has been called the *Monthly Infection Rate.* This depends chiefly on the number of infected anophelines in the neighbourhood, that is, the number which have bitten previously infected persons.

I shall in general follow the notation adopted by Professor Ross (*vide Report*, p. 31, *et seq.*).

Let p denote the average population of the locality during the enquiry, m the number of infected persons*, h the number of healthy persons at the beginning of the first month and im the average number of the infected persons whose blood contains enough of the malaria parasites to infect anophelines; evidently i is a fraction. Next, let a denote the average number of anophelines in the locality capable of carrying malaria and ba the average number of these which succeed in biting a single person during the month; here b must be a very small fraction,

* This differs from Professor Ross' use of m.

since probably many mosquitoes never succeed in biting men at all, and of those that do succeed, only a few will have bitten a particular individual. Hence the average number of anophelines which have bitten the suitably infected persons during the month will be *baim*.

Now if any of these insects are, in their turn, to infect human beings, they must survive for at least a week or ten days, in order to give time for the parasites to mature within them; and by no means all of them will survive so long. Let s be the average proportion of anophelines which can survive for that period; then *sbaim* will be the number of infected mosquitoes which have survived long enough to infect men in their turn, where s is also a fraction. But not all of these will find opportunities to bite human beings again, though they have survived long enough to do so. Let f be the proportion which succeed in biting. Then *fsbaim* will be the average number of infecting anophelines which succeed in biting men; also, if each of these bite a separate person and only one person, the same expression will denote the average number of persons bitten by infecting mosquitoes during the month. Call this number n. The most probable distribution of the n bites between malarial and non-malarial persons will be in the ratio $m : h$; so that the number of new infections during the month will be *fsbaim* h/p, or nh/p *.

Now each time a healthy person is bitten by an infecting mosquito the number of malarial patients is increased by unity, while the bites inflicted on those who are already suffering from the disease make no difference in the number of cases. This is equivalent to increasing m and decreasing h by h/p for each bite.

[By taking mean values in the above I am assuming that if

$$z = F(x_1, x_2, \dots x_n),$$

then $\bar{z}$ may be taken $= F(\bar{x}_1, \bar{x}_2, \dots \bar{x}_n)$, and I have been led to make this assumption in a *first* investigation by finding that the modal values of the distribution approximate very closely to the means in a number of particular cases fully worked out.]

Recovery Rate. The ratio of the number of recoveries to the total number of malaria cases during any period is called the *Recovery Rate* for that period. Professor Ross has estimated that of a given number of infected persons 50 °/₀ are ill after three months, that is, the recovery rate is ·5 per three months.

If m_0 be the number of cases at the beginning of the period, m_1, m_2, m_3, at the ends of the first, second and third months respectively, and r the recovery rate per month,

$$m_1 = m_0 - m_0 r = m_0(1 - r),$$
$$m_2 = m_1(1 - r) = m_0(1 - r)^2,$$
$$m_3 = m_2(1 - r) = m_0(1 - r)^3.$$

* Professor Ross assumes that to a first approximation n will be the number of new cases in the month.

But $$m_3 = m_0(1 - \cdot 5),$$

$$\therefore \quad (1-r)^3 = \cdot 5,$$

$$\log (1-r) = \tfrac{1}{3} \log \cdot 5$$

$$= \bar{1}\cdot 899657,$$

$$\therefore \quad 1 - r = \cdot 7937,$$

and $$\therefore \quad r = \cdot 2063 \text{ per month.}$$

In the numerical examples it will be necessary to know the recovery rate for the average period between two consecutive bites. Suppose there are n bites in a month and r is the recovery rate for the period between two consecutive bites, then, as before,

$$(1-r)^n = 1 - \cdot 2063$$

$$= \cdot 7937,$$

$$\log (1-r) = \frac{1}{n} \log \cdot 7937$$

$$= \frac{1}{n} \times \bar{1}\cdot 899657$$

$$= -\frac{\cdot 100343}{n},$$

whence the value of r is readily obtained.

I will now consider a formula for giving the number of cases at the end of a month in terms of the number at the beginning, the total population, the number of anophelines and the recovery rate.

Let m_0 be the number of cases at the beginning of the month out of a population p; m_t the number after the t^{th} bite and r the recovery rate when there are n infecting bites in the month; then

$$m_t = m_{t-1} + \frac{p - m_{t-1}}{p} - r m_{t-1},$$

or $$m_t = m_{t-1}\left(1 - \frac{1}{p} - r\right) + 1.$$

Put R for $\left(1 - \frac{1}{p} - r\right)$ and this equation may be written

$$m_t = m_{t-1} R + 1.$$

Now r remains constant for the month under consideration and therefore also R. Hence

$$m_{t-1} = m_{t-2} R + 1,$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$m_2 = m_1 R + 1,$$

$$m_1 = m_0 R + 1.$$

From these equations

$$m_t = m_0 R^t + \frac{1 - R^t}{1 - R},$$

and therefore, for the end of the month

$$m_n = m_0 R^n + \frac{1 - R^n}{1 - R} \quad \text{.....(A).}$$

Now, for any particular month $n = fsbaim$ where m is the number of cases at the beginning of the month, (or the average number for the first half of the month and the second half of the previous month); and Professor Ross has estimated the approximate values of the constants f, s, b, i to be $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{4p}$ and $\frac{1}{4}$ respectively; hence

$$n = \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{4p} \cdot \frac{1}{4} \cdot am$$

$$= \frac{am}{192p},$$

also $$r = \cdot 2063 \text{ when } n \not> 1,$$

and $$\log (1 - r) = - \frac{\cdot 100343}{n} \text{ when } n \not< 1.$$

We can thus by successive applications of the formula (A) obtain a series of values of m at the ends of consecutive months when we know the value at the beginning of the period, the total population and the average number of anophelines in the neighbourhood. Although it is impossible to estimate this last quantity with any degree of accuracy, yet by comparing the results of numerical examples, using different values of a, we get an idea of the effect on the malaria rate of reducing the anophelines in any given proportion.

Condition for a Stable Population. Suppose that a certain locality has a population p and m_0 cases of malaria. This latter will remain stationary if the number of new cases is equal to the number of recoveries. With a given value of m_0 and p the number of new infections depends on the number of anophelines in the neighbourhood. Hence with the anophelines present in a certain proportion to the human population, the malaria rate is stationary and if they are present in a greater or less proportion there will be a corresponding increase or decrease in the number of cases.

Let m_1 be the number of cases at the end of the month; then, by Equation (A)

$$m_1 = m_0 R^n + \frac{1 - R^n}{1 - R}.$$

The value of m is stationary if $m_1 = m_0 = m$, i.e.

$$m (1 - R^n) = \frac{1 - R^n}{1 - R},$$

or $$\left(m - \frac{1}{1 - R}\right)(1 - R^n) = 0,$$

$\therefore$ either $$1 - R^n = 0 \quad \text{.....(i)},$$

or $$m - \frac{1}{1 - R} = 0 \quad \text{.....(ii)}.$$

If $1-R^n=0$, $n=0$ for $R=1-\frac{1}{p}-r$ and is therefore fractional; also if $n=0$ the case reduces to that in which there is an entire absence of mosquitoes: hence in this enquiry $1-R^n\neq 0$ and we must proceed with Equation (ii). This may be written

$$r=\frac{1}{m}-\frac{1}{p}.$$

Now when $n>1$

$$\log(1-r)=-\frac{\cdot 100343}{n}.$$

Hence by giving a series of values to m we may find the number of anophelines corresponding to any stable population.

Taking $p=1000$, $m=5$, we get

$$r=\cdot 2-\cdot 001=\cdot 199,$$

$$\therefore\quad \log(1-r)=-\cdot 096367,$$

$$\therefore\quad \frac{\cdot 100343}{n}=\cdot 096367,$$

and

$$n=1\cdot 04125.$$

But

$$n=\frac{am}{192p},$$

$$\therefore\quad a=\frac{192\times 1000\times 1\cdot 04125}{5}$$

$$=39984.$$

Similarly we get the table of corresponding values of m and a for stable populations given below.

TABLE A.

Table of Corresponding Values of m and a for a Stable Population of a Thousand.

m	*a*	*m*	*a*	*m*	*a*
5	39984	*150*	52032	*600*	110914
8	41886	*175*	53648	*650*	126720
10	42553	*200*	55339	*700*	147892
15	43539	*250*	59062	*750*	177525
20	44148	*300*	63308	*800*	221752
25	44605	*350*	68193	*850*	295897
50	46251	*400*	73883	*900*	444118
75	47662	*450*	80612	*950*	885583
100	49069	*500*	88680		
125	50520	*550*	98534		

Again, suppose $m = 5$ and that a has such a value as will make $n = 1$, i.e.

$$\frac{5a}{192000} = 1 \text{ or } a = 38400.$$

We now have $r = \cdot 2063$ and $rm = 1 \cdot 0315$; hence $rm > n$; or the number of recoveries exceeds the number of new cases and the value of m decreases continuously.

From the curve connecting these quantities (Diagrams I and II), we see at a glance the value of m corresponding to any given value of a and we gather that any appreciable change in the number of anophelines will involve a corresponding gradual change in the number of cases; or, in other words, that, given any definite number of anophelines per thousand of the population there is a particular malaria rate which will ultimately be reached by a gradual change in the value of m.

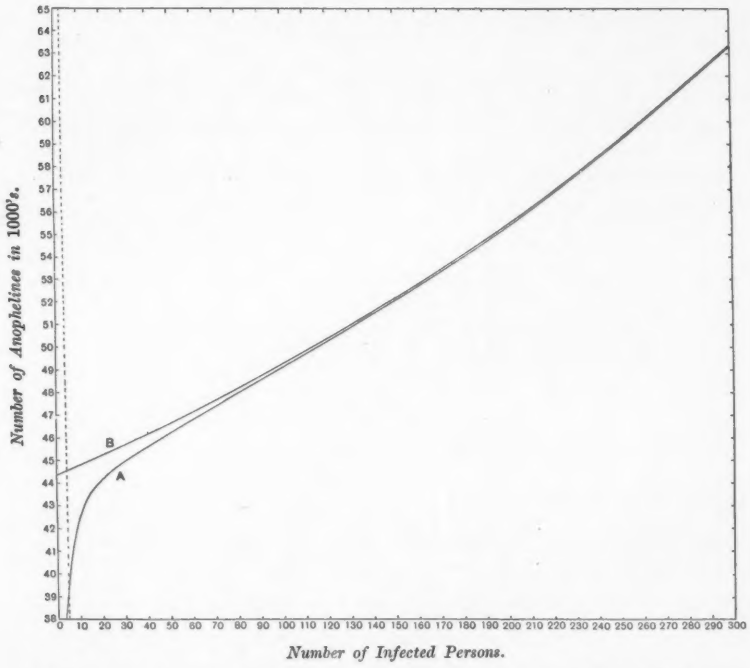


DIAGRAM I.* A. Stable Population Curve, Part 1. B. Curve of Equation (ii) p. 429.

* See p. 429, for the construction of these curves.

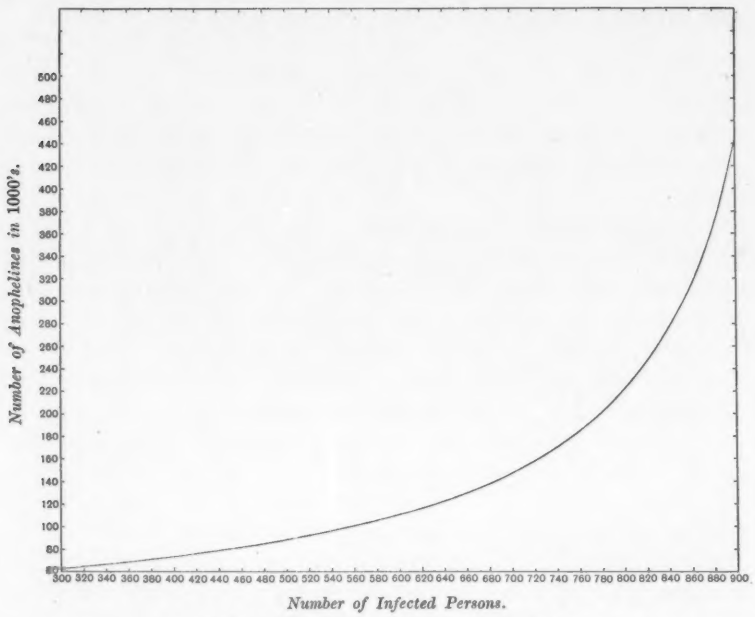


DIAGRAM II.* Stable Population Curve, Part 2.

The direct relation between m and a (these symbols denoting corresponding values for a stable population) may be investigated as follows:

When $n \nless 1$, we have (p. 424)

$$(1-r)^n = \cdot 7937,$$

$$\therefore \quad \log_e (1-r) = \frac{1}{n} \log_e \cdot 7937$$

$$= -\frac{\cdot 23104}{n},$$

or
$$1 - r = e^{-\frac{\cdot 23104}{n}}.$$

Now for a stable population

$$r = \frac{1}{m} - \frac{1}{p},$$

also
$$n = \frac{am}{192p},$$

$$\therefore \quad 1 + \frac{1}{p} - \frac{1}{m} = e^{-\frac{\cdot 23104 \times 192p}{am}}$$

$$= e^{-\frac{44 \cdot 36p}{am}}.$$

* See p. 429 for the construction of this curve.

Put c for 44·36 and the equation is*

$$1+\frac{1}{p}-\frac{1}{m}=e^{-\frac{pc}{am}} \quad \text{.....(i).}$$

Expanding the right-hand side we get, to a first approximation,

$$1+\frac{1}{p}-\frac{1}{m}=1-\frac{pc}{am},$$

which reduces to

$$a=\frac{p^2c}{p-m} \quad \text{.....(ii).}$$

A table of values of a and m found from Equation (ii) is given in Table B below; the graph is the upper curve in Diagram I, the portion of the graph common to this equation and to Table A, is given in Diagram II.

The graph of Equation (ii) fits very closely that of Table A when m exceeds 100, and accurately when m approaches p; for small values of m however the two

* The general form of this curve is given below. It has three branches :—(*a*) in the first quadrant, the line $m=p$ being an asymptote; this branch ends in a *point d'arrêt* at $\left(\frac{p}{p+1}, 0\right)$ on the m axis; (*b*) in the second quadrant with the axes of coordinates as asymptotes; (*c*) in the fourth quadrant, the axis of m and the line $m=p$ being asymptotes. That part of the branch (*a*) which is above the hyperbola $am=192p$, alone has any direct application to the present problem.

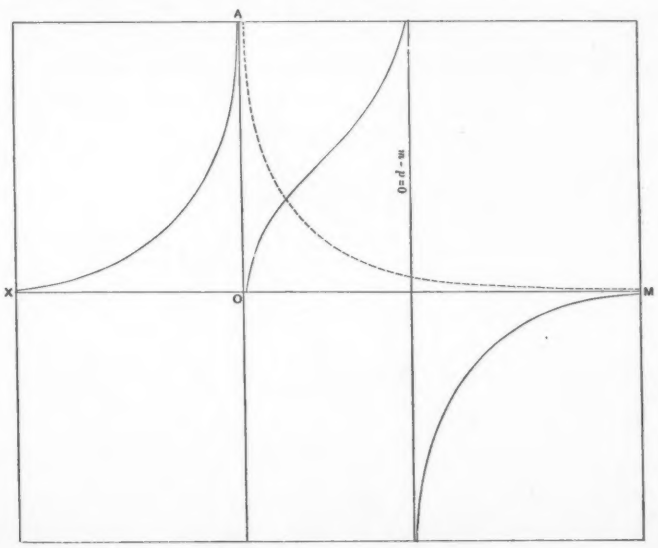


DIAGRAM III. General Form of Curve of Equation (i). The dotted line is the hyperbola which limits the practical solution of the part of the curve in the quadrant AOM.

curves differ considerably and it is necessary to proceed with Equation (i) to a higher degree of approximation.

TABLE B. (*Equation* (ii).)

m	a
0	44360
10	44808
50	46695
100	49288
150	52188
200	55450
250	59146
300	63371

Writing Equation (i) in the form

$$-\frac{pc}{am}=\log_e\left(1+\frac{1}{p}-\frac{1}{m}\right)$$

and expanding the right-hand side, we have

$$-\frac{pc}{am}=\left(\frac{1}{p}-\frac{1}{m}\right)-\frac{1}{2}\left(\frac{1}{p}-\frac{1}{m}\right)^2+\frac{1}{3}\left(\frac{1}{p}-\frac{1}{m}\right)^3-+\dots,$$

or

$$\frac{pc}{am}=\frac{p-m}{pm}+\frac{1}{2}\frac{(p-m)^2}{p^2m^2}+\frac{1}{3}\frac{(p-m)^3}{p^3m^3}+\dots.$$

Putting z for $\frac{p-m}{pm}$, this may be written:

$$\frac{pc}{am}=z+\frac{1}{2}z^2+\frac{1}{3}z^3 \text{ approx.},$$

or

$$\begin{aligned}\frac{am}{pc}&=\frac{1}{z}\left(1+\frac{1}{2}z+\frac{1}{3}z^2\right)^{-1}\\&=\frac{1}{z}\left\{1-\frac{1}{2}z-\frac{1}{3}z^2+\left(\frac{1}{2}z+\frac{1}{3}z^2\right)^2-+\dots\right\}\\&=\frac{1}{z}\left\{1-\frac{1}{2}z-\frac{1}{3}z^2+\frac{1}{4}z^3+\dots\right\}\\&=\frac{1}{z}-\frac{1}{2}-\frac{z}{12} \text{ approx.},\end{aligned}$$

i.e.

$$a=\frac{p^2c}{p-m}-\frac{pc}{2m}-\frac{c(p-m)}{12m^2} \qquad \text{.....(iii).}$$

The above relations cease to hold good when $n<1$, i.e. when $am<192p$; hence the limit of applicability of the *Stable Population Curve*, obtained from Equation (iii), is given by its intersection with the hyperbola

$$am=192p.$$

Part of this hyperbola is shown on the left of Diagram I as a broken line from which it appears that in a population of a thousand, there can be no stable condition with a smaller average number of infected persons than about 4·5.

The numbers in Table A are all found for a population of a thousand, but the general form of Equation (iii) is applicable to a population of any magnitude so long as the number of infecting bites does not fall below one per month. For example, when $p = 20000$, and $m = 200$, Equation (iii) gives $a = 893941$, while the formulae on p. 426 give $a = 893966$, the difference being less than ·003 per cent.

It will be of interest to compare the values of a found from Equation (iii) with those of Table A in a few typical cases:

m	a, by Equation (iii)	a, from Table A
10	42554	42553
50	46249	46251
100	49067	49069
500	88676	88680

In the last three cases the term $\frac{c(p-m)}{12m^2}$ is negligible, while in the last case the term $\frac{pc}{2m}$ is only about ·01 per cent. of the whole.

Numerical Examples. In the numerical illustrations which follow, the increase or decrease in malaria, month by month, has been obtained by successive applications of Equation (A), p. 425, in the form

$$m_s = m_{s-1}R_s^{n_s} + \frac{1 - R_s^{n_s}}{1 - R_s},$$

where

$$n_s = \frac{am_{s-1}}{192p};$$

$$R_s = 1 - \frac{1}{p} - r_s,$$

$$\log(1 - r_s) = -\frac{\cdot 100343}{n_s} \qquad [n > 1],$$

$$r_s = \cdot 2063 \qquad [n \not> 1].$$

Seven-figure tables have been used throughout in the calculations but the results are given to one place of decimals only.

[*Note.* It is assumed, in all the examples, that Professor Ross' recovery rate holds. This would rapidly reduce a large number of cases to the vanishing point in the absence of new infections, whereas it is well known that a few persons suffer from relapses after living many years in countries where there is no possibility of re-infection. (See *Report*, p. 16.)]

Example 1, Case (i). $p = 1000$, $m_0 = 50$, $a = 64000$. (The number of anophelines for a stable population is about 46000.)

TABLE C.

(*Giving the value of m for consecutive months.*)

Months	1st Year	2nd Year	3rd Year	4th Year	5th Year	6th Year
1	53·9	118·0	198·8	258·5	286·5	297·7
2	58·0	124·6	205·0	261·7	287·9	298·2
3	62·4	131·3	211·0	264·9	289·2	298·5
4	67·0	138·0	216·8	267·8	290·3	299·0
5	71·7	144·9	222·4	270·5	291·4	299·4
6	76·7	151·8	227·7	273·1	292·4	299·6
7	82·0	158·6	233·0	275·5	293·3	300·1
8	87·5	165·5	238·0	277·4	294·3	300·4
9	93·2	172·6	242·6	279·5	295·0	300·6
10	99·1	179·5	247·0	281·5	295·7	301·0
11	105·2	186·1	251·2	283·3	296·5	
12	111·6	192·6	254·9	284·9	297·1	

Example 1, Case (ii). $p = 1000$, $m_0 = 50$, $a = 24000$.

TABLE D.

Months	1st Year	2nd Year	3rd Year	4th Year
1	45·1	13·8	5·8	4·98
2	40·7	12·6	5·7	4·96
3	36·8	11·5	5·5	4·94
4	33·2	10·5	5·4	4·92
5	30·1	9·6	5·3	4·91
6	27·2	8·8	5·3	4·90
7	24·7	8·0	5·2	4·89
8	22·4	7·4	5·1	4·88
9	20·3	6·9	5·1	4·87
10	18·3	6·5	5·1	4·87
11	16·6	6·2	5·0	4·86
12	15·2	6·0	5·0	4·86

It will be seen from this example that with a comparatively small number of infected persons and a quantity of anophelines somewhat in excess of the number which would keep the population stable, we get a steady rise in the malaria rate until nearly one-third of the whole population are infected; on the other hand, the destruction of rather more than half the mosquitoes will produce a fall in the rate until the cases are so few that they can be easily and effectively dealt with.

For the graphs see Diagram IV.

Example 2. $p = 1000$, $m_0 = 500$, $a =$ (i) 96000, (ii) 48000, (iii) 24000, (iv) 0. (The number of anophelines corresponding to a stable population would be about 89600.)

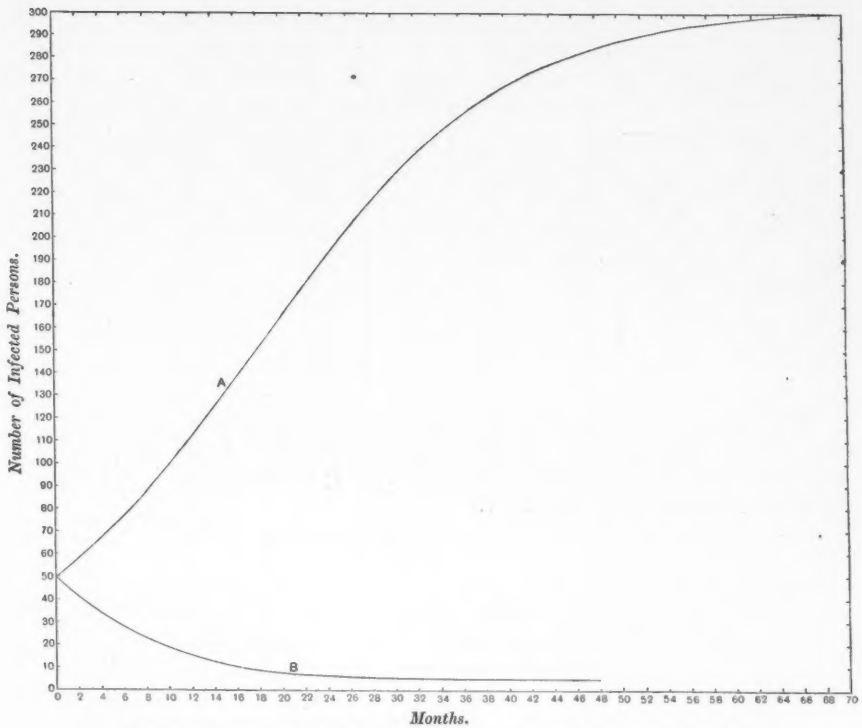


DIAGRAM IV. Curves illustrating Example 1. A. Case (i). B. Case (ii).

In Case (i) we get an increase and in each of the other cases a decrease in the value of m according to the following Tables:

TABLE E. *Case* (i), $a=96000$.

TABLE F. *Case* (ii), $a=48000$.

Months	1st Year	1st Year	2nd Year	3rd Year	4th Year	5th Year
1	507·9	455·3	235·0	169·7	138·9	121·3
2	513·9	418·7	227·0	166·3	137·0	120·2
3	518·7	388·3	219·6	163·1	135·3	119·1
4	522·7	362·4	212·8	160·2	133·6	118·0
5	526·3	340·2	206·5	157·3	132·0	117·0
6	528·8	321·0	200·7	154·6	130·5	116·0
7	530·7	304·1	195·3	152·0	129·0	115·1
8	532·3	289·3	190·3	149·6	127·6	114·2
9	533·5	276·0	185·6	147·2	126·2	113·3
10	534·4	264·2	181·3	145·0	124·9	112·5
11	535·2	253·5	177·2	142·9	123·6	111·7
12	535·8	243·9	173·3	140·8	122·4	110·9

TABLE G. *Case* (iii), $a = 24000$.

TABLE H. *Case* (iv), $a = 0$.

Months	1st Year	2nd Year	3rd Year	4th Year	1st Year	2nd Year	3rd Year
1	427·0	94·8	27·3	8·8	396·9	24·8	1·6
2	369·8	85·0	24·7	8·1	315·0	19·7	
3	320·8	76·2	22·4	—	250·0	15·6	
4	279·8	68·5	20·3	—	198·4	12·4	
5	245·3	61·6	18·5	—	157·5	9·8	
6	215·8	55·5	16·8	—	125·0	7·8	
7	190·6	50·0	15·3	—	99·2	6·2	
8	168·8	45·2	13·9	—	78·7	4·9	
9	149·9	40·8	12·7	—	62·5	3·9	
10	132·9	36·8	11·6	—	49·6	3·1	
11	118·5	33·3	10·5	—	39·4	2·5	
12	105·9	30·1	9·6	—	31·3	1·9	

This example serves to show the effect (i) of leaving the mosquitoes alone, (ii) of destroying half, (iii) of destroying three-quarters of the mosquitoes, (iv) of exterminating them (but see *Note* on p. 431); the comparison is best seen by means of the curves in Diagram V.

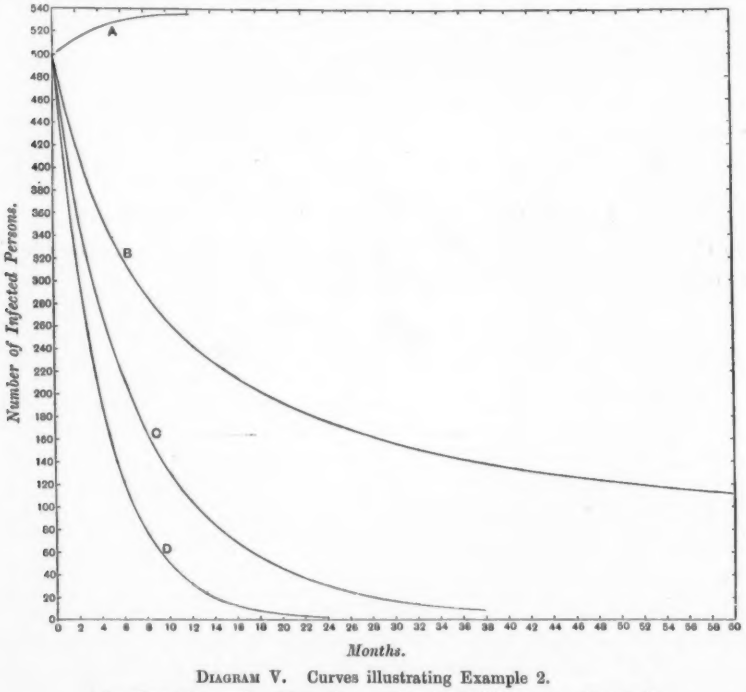


Diagram V. Curves illustrating Example 2.
A. Case (i). B. Case (ii). C. Case (iii). D. Case (iv).

The foregoing results point to a stable population for any given number of mosquitoes, but it is perfectly clear that this number is itself subject to constant change, often of considerable magnitude, and that, in consequence, a stable population is rarely, if ever, attained in practice.

The minor contributory causes mentioned on p. 421, will also affect the malaria rate to an extent which cannot be neglected, while another important item to consider is the distribution both of the human population and of the mosquitoes; also whether the most densely populated districts are at the same time those most infected with mosquitoes. These and many other points will require careful investigation with the aid to be derived from the most recent and complete observations before the subject can be said to be exhausted, and I am quite well aware that the present paper merely traverses the outskirts of an important and extensive field of research.

But still it seems to me to establish two very important points, i.e., (i) Given a number of mosquitoes greater than that corresponding to the "stable population value," the number of malaria cases will tend to increase until a stable population value is reached, (ii) Given a number of mosquitoes less than that corresponding to the stable population value, the number of malaria cases will tend to decrease until that stable population is established.

In both cases the number of malaria cases tends to asymptote to the stable value. The amount of malaria does not increase or decrease indefinitely, but tends to attain a definite prevalence. Where the "stable value" means a large number of malaria cases, the right step seems to be the reduction of the number of mosquitoes; on the other hand, where it means a small number of cases it should be possible to segregate and isolate these cases.

In conclusion I wish to express my grateful thanks to Professor Karl Pearson for much valuable assistance in the preparation of this paper.

[*Postscript.* Since completing this paper, I have, through the kindness of Professor Ross, been able to see proofs of part of his forthcoming treatise on Malaria. From a comparison of the two accounts it will be seen that there is complete agreement between us on the essential and fundamental points, but some difference in numerical details. This difference is due to an attempt on my part to treat the question by fuller and more rigid mathematical methods which would be out of place in a treatise written from the medical standpoint, while the simpler methods employed by Professor Ross give results which are sufficiently accurate for practical purposes.

The principal points of agreement are: (*a*) for a given number of anophelines per unit of the population the number of malaria cases will gradually rise or fall to a fixed value at which it will remain stationary, and (*b*) when the anophelines are less than a certain number (about forty per unit of the population) there can be

no stable condition and the malaria cases will gradually decrease and finally disappear.

The differences are: (*c*) the number of cases in the "stable population," corresponding to a given number of anophelines, is higher in Professor Ross' treatment than in mine, and (*d*) the changes in the malaria rate are more rapid; e.g., in a particular example, where the population is 1000, the number of anophelines per person 100 and the initial number of cases 500, we have:

Months.....	0	1	2	3	4...	finally
Cases (Ross)	500	525	544	560	571	600
„ (Waite)	500	512	521	529	535	556

The divergence seems to be chiefly due to the difference in the time units employed in the two methods of treatment. Professor Ross has used the *month* throughout and has taken the value of m constant during each month, while I have used the average time between two consecutive infecting bites as my unit. The fact that m is increased by unity each time a healthy person is bitten by an infected mosquito and is continually being diminished owing to recoveries, fully justifies, in my opinion, the adoption of this unit.

Part of the difference (*d*) is easily accounted for as follows: Suppose m is increasing; then in finding m_{s+1} from m_s, the number of new cases obtained by Professor Ross is too great; for m_s is the smallest value of m during the month, and therefore the chance of malarial sufferers being bitten again is greater and that of healthy persons being infected less than that found. Again, the number of recoveries found for the month is, for the same reason, too small, and thus both causes combine to make the value of m_{s+1} too great.

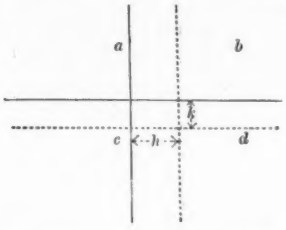
Similar reasoning shows that when the malaria is decreasing the value found for m_{s+1} is too small; i.e. in both cases the changes in the malaria rate are too rapid. It is evident, however, that the difference on this account becomes smaller as the stable value is approached. H. W.]

TABLES OF THE TETRACHORIC FUNCTIONS FOR FOURFOLD CORRELATION TABLES.

By P. F. EVERITT, B.Sc.

Explanation of the Tables.

The method of the fourfold table for determining correlation was described by Pearson in *Phil. Trans.* A. vol. 195, pp. 1—47.



a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	N

Following his nomenclature, the normal correlation surface is divided into four parts by two planes at right angles to the axes of x and y at distances h', k' from the origin and these distances h' and k', when measured in terms of the standard deviations of their respective variables, are called h and k. The volumes or frequencies in the four divisions are represented by a, b, c, d in the manner shown in the plan and it will be seen, that $b+d$ and $c+d$, owing to the position given to the point of intersection of the traces of the dividing planes, cannot exceed $\frac{1}{2}N$.

In the paper referred to, the correlation coefficient r is determined from the equation

$$\frac{d}{N} = \frac{b+d}{N} \cdot \frac{c+d}{N} + \mathop{S}_{1}^{\infty}\left(\frac{r^n}{n!} HK\, \bar{v}_{n-1}\, \overline{w}_{n-1}\right),$$

where H, K are the ordinates of the normal curve of area N corresponding to the abscissae h and k and consequently dividing the curve into areas, of which the proportions to the whole are $\frac{b+d}{N}$ and $\frac{c+d}{N}$ respectively, while $\bar{v}$ and $\overline{w}$ are given by

$$\bar{v}_n = h\bar{v}_{n-1} - (n-1)\,\bar{v}_{n-2}, \qquad \overline{w}_n = k\overline{w}_{n-1} - (n-1)\,\overline{w}_{n-2},$$
$$\bar{v}_0 = 1, \quad \bar{v}_1 = h, \qquad \overline{w}_0 = 1, \quad \overline{w}_1 = k.$$

Now let

$$\tau_n = \frac{H\bar{v}_{n-1}}{\sqrt{n!}} \text{ and } \tau_n' = \frac{K\bar{w}_{n-1}}{\sqrt{n!}},$$

then the equation becomes

$$\frac{d}{N} = \frac{b+d}{N} \cdot \frac{c+d}{N} + \sum_1^\infty (\tau_n \tau_n' r^n).$$

It is clear from the above, that τ_n' is the same function of $\frac{c+d}{N}$ as τ_n is of $\frac{b+d}{N}$ and that one table of these functions will serve for both, if we enter the table with $\frac{b+d}{N}$ as argument for the latter, and with $\frac{c+d}{N}$ for the former. It should be noted that these quantities $\frac{b+d}{N}$, $\frac{c+d}{N}$ are identical with $\frac{1}{2}(1-\alpha)$, where $\frac{1}{2}(1+\alpha)$ and α are used as arguments by Sheppard* in his published tables and to avoid ambiguity have been tabulated under that heading.

In the present tables the values of the first six τ functions, henceforth to be termed *tetrachoric* functions, have been computed for values of $\frac{1}{2}(1-\alpha)$ from ·001 to ·500 by successive increments of ·001; the last column contains the values of h† (or k) corresponding to the value of $\frac{1}{2}(1-\alpha)$ given in the first column, and required for computing the functions of higher order than the sixth as well as for the probable error of r.

In the auxiliary table are given the values of p_n and q_n required to compute the functions of the seventh to twelfth orders by means of the difference formula

$$\tau_n = hp_n\tau_{n-1} - q_n\tau_{n-2}.$$

Illustrations of the Use of the Tables.

(*a*) With interpolation.

Consider the hypothetical table given below:

1668	131	1799
137	64	201
1805	195	2000

Here $\frac{b+d}{N} = ·09750$, $\frac{c+d}{N} = ·10050$, $\frac{d}{N} = ·032000$.

* W. F. Sheppard, "New Tables of the Probability Integral," *Biometrika*, Vol. II. p. 174.
† x of Sheppard's Tables.

Now enter the table twice using interpolation, once with $\frac{1}{2}(1-\alpha)$ equal to ·09750 and a second time with it equal to ·10050, and we have

$\frac{1}{2}(1-\alpha)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
·09750	+ ·17228	+ ·15787	+ ·04779	− ·06018	− ·06693	+ ·00854
·10050	+ ·17614	+ ·15926	+ ·04567	− ·06275	− ·06652	+ ·01111

Multiplying the numbers in each column together, we find the equation

$$·032000 = ·009799 + ·030345r + ·025142r^2 + ·002183r^3 + ·003776r^4 + ·004452r^5 + ·000095r^6;$$

whence, solving by Newton's method,

$$r = ·501.$$

An illustration of the calculation of the probable error of r when found as above is given in Pearson's paper on p. 36.

(*b*) Without interpolation.

Using the same table as in example (*a*), we have as before

$$\frac{b+d}{N} = ·09750, \quad \frac{c+d}{N} = ·10050, \quad \frac{d}{N} = ·032000.$$

Entering the table with $\frac{1}{2}(1-\alpha)$ equal to ·098 and ·101 we have

$\frac{1}{2}(1-\alpha)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
·098	+ ·17292	+ ·15811	+ ·04744	− ·06061	− ·06687	+ ·00897
·101	+ ·17678	+ ·15948	+ ·04531	− ·06317	− ·06644	+ ·01153

Multiplying out as before

$$·032000 = ·009898 + ·030569r + ·025215r^2 + ·002150r^3 + ·003829r^4 + ·004443r^5 + ·000103r^6;$$

whence, solving by Newton's method,

$$r = ·498.$$

This particular illustration, chosen so that the true values of $\frac{1}{2}(1-\alpha)$ fall midway between tabulated values, shows the maximum error caused by working without interpolation for values of $\frac{1}{2}(1-\alpha)$ of ·100, and in practically every case this error will be negligible when compared with the probable error of r.

A study of the differences of the functions shows that for values of $\frac{1}{2}(1-\alpha)$ greater than ·100 the error introduced by working without interpolation will not be appreciably greater than in the example given and may quite easily be less; for values of $\frac{1}{2}(1-\alpha)$ less than ·100 it will be desirable to use interpolation if the greatest accuracy attainable is desired, but even in very unfavourable cases such errors will rarely become as large as the probable error of the result.

The tetrachoric functions of higher order than the sixth (easily found by means of the auxiliary table) will be chiefly required when r is large and the series consequently converges less rapidly. In connection with such cases, it is well to recollect that in the series on the right-hand side of the equation successive terms need not necessarily converge but that *alternate* terms are absolutely convergent*.

Another Use for the Tables.

The tables can also be used with advantage in the work of computing the frequencies in the divisions of any frequency table, when the correlation coefficient r is known and the frequency is normal. Such problems can be classified into three cases according to the fineness of the grouping and a brief outline of the methods to be employed is given.

Case I. Fourfold table. According to the purpose to be served, the dividing planes are fixed either by h and k or $b+d$ and $c+d$ being known; the value of d is then calculated from

$$\frac{d}{N} = \frac{b+d}{N} \cdot \frac{c+d}{N} + \sum_{1}^{\infty} (\tau_n \tau_n' r^n);$$

when d is known, b, c and a are also known from the total column and row.

Case II. Ninefold table. It is required to divide a normal frequency surface into nine parts A, B,... L as shown in the diagram. Then it will always be possible

A	*B*	*C*	
D	*E*	*F*	
G	*I*	*L*	
			N

to consider one or more of A, C, G, L in turn as the d of a standard fourfold table and find its value from the equation given in Case I. The column and row totals are either known or obtained directly from values of h and k. From the known divisions and the column and row totals the values of other divisions and groups of divisions are next obtained; the table is then divided afresh, if necessary, according to the particular circumstances of the case, the values of other groups of divisions obtained, and so ultimately the value of each separate division not already known may be found by differences.

* Pearson, *loc. cit.* p. 10.

Case III. More than three columns or rows.

Find the value of as many of the corner divisions as may be treated as the d of a fourfold table by the method of Case I; then find the column and row totals, if not already known, and from them the values of other divisions and groups of divisions. The table is then divided afresh and the values of other groups of divisions obtained and this process repeated until the values of the remaining separate divisions may be found by differences. This case is really an extension of Case II.

In conclusion I desire to express my indebtedness to Prof. Pearson for suggesting the construction of the tables and for much valuable advice as to methods of calculation and arrangement. I also wish to express my thanks to Mr Sheppard for allowing me to use a hitherto unpublished table of the probability integral, from which the values of h corresponding to given values of $\frac{1}{2}(1-\alpha)$ have been taken, the published tables not providing sufficient decimal places without troublesome interpolation.

Auxiliary Table for calculating Tetrachoric Functions beyond those Tabled.

n	p_n	q_n
7	·37796	·77152
8	·35355	·80178
9	·33333	·82496
10	·31623	·84327
11	·30151	·85812
12	·28868	·87039

$$\tau_n = hp_n\tau_{n-1} - q_n\tau_{n-2}.$$

TABLE OF THE TETRACHORIC FUNCTIONS.

$\frac{1}{2}(1-a)$	r_1	r_2	r_3	r_4	r_5	r_6	h
·001	·00337	·00736	+ ·01175	+ ·01391	+ ·01134	+ ·00415	3·09023
·002	·00634	·01290	+ ·01885	+ ·01968	+ ·01269	+ ·00053	2·87816
·003	·00915	·01778	+ ·02446	+ ·02335	+ ·01228	− ·00328	2·74778
·004	·01185	·02222	+ ·02918	+ ·02587	+ ·01111	− ·00687	2·65207
·005	·01446	·02634	+ ·03326	+ ·02764	+ ·00952	− ·01017	2·57583
·006	·01700	·03020	+ ·03686	+ ·02887	+ ·00770	− ·01318	2·51214
·007	·01949	·03386	+ ·04008	+ ·02970	+ ·00575	− ·01592	2·45726
·008	·02192	·03734	+ ·04298	+ ·03021	+ ·00371	− ·01841	2·40892
·009	·02431	·04066	+ ·04561	+ ·03047	+ ·00164	− ·02067	2·36562
·010	·02665	·04384	+ ·04800	+ ·03053	− ·00044	− ·02271	2·32635
·011	·02896	·04690	+ ·05020	+ ·03041	− ·00253	− ·02457	2·29037
·012	·03123	·04985	+ ·05221	+ ·03014	− ·00460	− ·02625	2·25713
·013	·03348	·05270	+ ·05406	+ ·02975	− ·00664	− ·02777	2·22621
·014	·03569	·05545	+ ·05577	+ ·02926	− ·00866	− ·02914	2·19729
·015	·03787	·05811	+ ·05735	+ ·02867	− ·01064	− ·03037	2·17009
·016	·04003	·06069	+ ·05880	+ ·02801	− ·01259	− ·03147	2·14441
·017	·04216	·06320	+ ·06015	+ ·02727	− ·01449	− ·03246	2·12007
·018	·04427	·06564	+ ·06139	+ ·02647	− ·01636	− ·03334	2·09693
·019	·04635	·06801	+ ·06254	+ ·02562	− ·01818	− ·03411	2·07485
·020	·04842	·07031	+ ·06361	+ ·02472	− ·01996	− ·03479	2·05375
·021	·05046	·07256	+ ·06459	+ ·02378	− ·02170	− ·03538	2·03352
·022	·05249	·07475	+ ·06549	+ ·02280	− ·02340	− ·03589	2·01409
·023	·05449	·07688	+ ·06633	+ ·02179	− ·02505	− ·03632	1·99539
·024	·05648	·07897	+ ·06709	+ ·02074	− ·02666	− ·03667	1·97737
·025	·05845	·08100	+ ·06780	+ ·01968	− ·02823	− ·03696	1·95996
·026	·06040	·08299	+ ·06844	+ ·01858	− ·02976	− ·03718	1·94313
·027	·06233	·08493	+ ·06903	+ ·01747	− ·03125	− ·03734	1·92684
·028	·06425	·08682	+ ·06956	+ ·01634	− ·03270	− ·03744	1·91104
·029	·06615	·08868	+ ·07005	+ ·01520	− ·03411	− ·03749	1·89570
·030	·06804	·09049	+ ·07048	+ ·01404	− ·03547	− ·03749	1·88079
·031	·06992	·09227	+ ·07087	+ ·01287	− ·03680	− ·03744	1·86630
·032	·07177	·09400	+ ·07122	+ ·01168	− ·03810	− ·03734	1·85218
·033	·07362	·09570	+ ·07153	+ ·01049	− ·03935	− ·03720	1·83842
·034	·07545	·09737	+ ·07179	+ ·00929	− ·04057	− ·03702	1·82501
·035	·07727	·09900	+ ·07202	+ ·00809	− ·04176	− ·03680	1·81191
·036	·07908	·10060	+ ·07221	+ ·00688	− ·04291	− ·03654	1·79912
·037	·08087	·10216	+ ·07237	+ ·00566	− ·04402	− ·03624	1·78661
·038	·08265	·10370	+ ·07249	+ ·00444	− ·04510	− ·03592	1·77438
·039	·08442	·10520	+ ·07258	+ ·00322	− ·04615	− ·03556	1·76241
·040	·08617	·10668	+ ·07264	+ ·00200	− ·04717	− ·03517	1·75069
·041	·08792	·10812	+ ·07268	+ ·00077	− ·04815	− ·03475	1·73920
·042	·08965	·10954	+ ·07268	− ·00045	− ·04910	− ·03431	1·72793
·043	·09137	·11093	+ ·07266	− ·00167	− ·05003	− ·03384	1·71689
·044	·09309	·11229	+ ·07261	− ·00290	− ·05092	− ·03335	1·70604
·045	·09479	·11363	+ ·07253	− ·00412	− ·05178	− ·03283	1·69540
·046	·09648	·11495	+ ·07243	− ·00534	− ·05261	− ·03229	1·68494
·047	·09816	·11623	+ ·07231	− ·00656	− ·05342	− ·03173	1·67466
·048	·09983	·11750	+ ·07217	− ·00778	− ·05420	− ·03115	1·66456
·049	·10149	·11874	+ ·07200	− ·00899	− ·05495	− ·03055	1·65463
·050	·10314	·11996	+ ·07181	− ·01020	− ·05567	− ·02994	1·64485

r_1, r_2 and h are essentially positive.

TABLE—(*continued*).

$\frac{1}{2}(1-a)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	h
·051	·10478	·12115	+·07160	−·01140	−·05637	−·02931	1·63523
·052	·10641	·12232	+·07138	−·01260	−·05704	−·02866	1·62576
·053	·10803	·12347	+·07113	−·01380	−·05769	−·02799	1·61644
·054	·10964	·12460	+·07087	−·01499	−·05831	−·02732	1·60725
·055	·11124	·12571	+·07058	−·01618	−·05891	−·02662	1·59819
·056	·11284	·12680	+·07028	−·01736	−·05949	−·02592	1·58927
·057	·11442	·12787	+·06997	−·01854	−·06004	−·02520	1·58047
·058	·11600	·12892	+·06964	−·01971	−·06057	−·02447	1·57179
·059	·11756	·12995	+·06929	−·02087	−·06107	−·02373	1·56322
·060	·11912	·13096	+·06893	−·02203	−·06155	−·02298	1·55477
·061	·12067	·13196	+·06855	−·02318	−·06202	−·02222	1·54643
·062	·12222	·13293	+·06816	−·02433	−·06246	−·02145	1·53820
·063	·12375	·13389	+·06775	−·02547	−·06288	−·02068	1·53007
·064	·12528	·13483	+·06734	−·02660	−·06328	−·01989	1·52204
·065	·12679	·13575	+·06690	−·02773	−·06365	−·01910	1·51410
·066	·12830	·13666	+·06646	−·02884	−·06401	−·01830	1·50626
·067	·12981	·13754	+·06601	−·02996	−·06435	−·01749	1·49851
·068	·13130	·13842	+·06554	−·03106	−·06467	−·01668	1·49085
·069	·13279	·13927	+·06506	−·03216	−·06498	−·01586	1·48328
·070	·13427	·14011	+·06457	−·03325	−·06526	−·01504	1·47579
·071	·13574	·14094	+·06407	−·03433	−·06552	−·01421	1·46838
·072	·13720	·14175	+·06356	−·03541	−·06577	−·01337	1·46106
·073	·13866	·14254	+·06304	−·03648	−·06600	−·01253	1·45381
·074	·14011	·14332	+·06251	−·03754	−·06621	−·01169	1·44663
·075	·14156	·14409	+·06197	−·03859	−·06641	−·01085	1·43953
·076	·14299	·14484	+·06142	−·03963	−·06659	−·01000	1·43250
·077	·14442	·14558	+·06086	−·04067	−·06675	−·00915	1·42554
·078	·14584	·14630	+·06029	−·04170	−·06690	−·00829	1·41865
·079	·14726	·14701	+·05971	−·04272	−·06703	−·00743	1·41183
·080	·14867	·14771	+·05913	−·04374	−·06715	−·00658	1·40507
·081	·15007	·14839	+·05854	−·04474	−·06725	−·00572	1·39838
·082	·15146	·14906	+·05794	−·04574	−·06733	−·00485	1·39174
·083	·15285	·14971	+·05733	−·04673	−·06741	−·00399	1·38517
·084	·15423	·15036	+·05671	−·04771	−·06746	−·00312	1·37866
·085	·15561	·15099	+·05609	−·04869	−·06751	−·00226	1·37220
·086	·15698	·15160	+·05546	−·04965	−·06753	−·00139	1·36581
·087	·15834	·15221	+·05483	−·05061	−·06755	−·00053	1·35946
·088	·15970	·15280	+·05418	−·05156	−·06755	+·00034	1·35317
·089	·16105	·15339	+·05353	−·05250	−·06754	+·00120	1·34694
·090	·16239	·15396	+·05288	−·05344	−·06751	+·00207	1·34076
·091	·16373	·15451	+·05222	−·05436	−·06748	+·00294	1·33462
·092	·16506	·15506	+·05155	−·05528	−·06743	+·00380	1·32854
·093	·16639	·15560	+·05088	−·05619	−·06736	+·00467	1·32251
·094	·16770	·15612	+·05020	−·05709	−·06729	+·00553	1·31652
·095	·16902	·15663	+·04952	−·05798	−·06720	+·00639	1·31058
·096	·17033	·15713	+·04883	−·05887	−·06710	+·00725	1·30469
·097	·17163	·15763	+·04813	−·05975	−·06699	+·00811	1·29884
·098	·17292	·15811	+·04744	−·06061	−·06687	+·00897	1·29303
·099	·17421	·15858	+·04673	−·06148	−·06674	+·00982	1·28727
·100	·17550	·15904	+·04602	−·06233	−·06660	+·01068	1·28155

TABLE—(*continued*).

$\frac{1}{2}(1-a)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	h
·101	·17678	·15948	+·04531	−·06317	−·06644	+·01153	1·27587
·102	·17805	·15992	+·04459	−·06401	−·06628	+·01238	1·27024
·103	·17932	·16035	+·04387	−·06484	−·06610	+·01322	1·26464
·104	·18058	·16077	+·04315	−·06566	−·06592	+·01407	1·25908
·105	·18184	·16118	+·04242	−·06647	−·06572	+·01491	1·25357
·106	·18309	·16158	+·04169	−·06727	−·06551	+·01575	1·24808
·107	·18433	·16197	+·04095	−·06807	−·06530	+·01659	1·24264
·108	·18557	·16235	+·04021	−·06886	−·06507	+·01742	1·23723
·109	·18681	·16272	+·03947	−·06964	−·06484	+·01825	1·23186
·110	·18804	·16308	+·03872	−·07041	−·06459	+·01908	1·22653
·111	·18926	·16343	+·03797	−·07117	−·06434	+·01990	1·22123
·112	·19048	·16378	+·03721	−·07193	−·06408	+·02072	1·21596
·113	·19169	·16411	+·03646	−·07268	−·06381	+·02154	1·21073
·114	·19290	·16443	+·03570	−·07342	−·06353	+·02235	1·20553
·115	·19410	·16475	+·03493	−·07415	−·06324	+·02316	1·20036
·116	·19530	·16506	+·03417	−·07488	−·06294	+·02397	1·19522
·117	·19649	·16536	+·03340	−·07559	−·06264	+·02477	1·19012
·118	·19768	·16565	+·03263	−·07630	−·06233	+·02557	1·18504
·119	·19886	·16593	+·03186	−·07700	−·06201	+·02636	1·18000
·120	·20004	·16620	+·03108	−·07770	−·06168	+·02716	1·17499
·121	·20121	·16647	+·03030	−·07838	−·06134	+·02794	1·17000
·122	·20238	·16672	+·02952	−·07906	−·06100	+·02873	1·16505
·123	·20354	·16697	+·02874	−·07973	−·06065	+·02950	1·16012
·124	·20470	·16721	+·02796	−·08039	−·06029	+·03028	1·15522
·125	·20585	·16745	+·02717	−·08105	−·05992	+·03105	1·15035
·126	·20700	·16767	+·02638	−·08169	−·05955	+·03181	1·14551
·127	·20814	·16789	+·02559	−·08233	−·05917	+·03257	1·14069
·128	·20928	·16810	+·02480	−·08297	−·05878	+·03333	1·13590
·129	·21042	·16830	+·02401	−·08359	−·05839	+·03408	1·13113
·130	·21155	·16849	+·02321	−·08421	−·05799	+·03483	1·12639
·131	·21267	·16868	+·02241	−·08482	−·05758	+·03557	1·12168
·132	·21379	·16886	+·02162	−·08542	−·05717	+·03631	1·11699
·133	·21490	·16903	+·02082	−·08601	−·05675	+·03704	1·11232
·134	·21601	·16919	+·02001	−·08660	−·05632	+·03777	1·10768
·135	·21712	·16935	+·01921	−·08718	−·05589	+·03850	1·10306
·136	·21822	·16950	+·01841	−·08775	−·05546	+·03921	1·09847
·137	·21932	·16964	+·01760	−·08831	−·05501	+·03993	1·09390
·138	·22041	·16978	+·01680	−·08887	−·05456	+·04064	1·08935
·139	·22149	·16990	+·01599	−·08942	−·05411	+·04134	1·08482
·140	·22258	·17003	+·01518	−·08996	−·05365	+·04204	1·08032
·141	·22365	·17014	+·01437	−·09050	−·05318	+·04273	1·07584
·142	·22473	·17025	+·01356	−·09103	−·05271	+·04342	1·07138
·143	·22580	·17035	+·01275	−·09155	−·05224	+·04410	1·06694
·144	·22686	·17044	+·01194	−·09206	−·05176	+·04478	1·06252
·145	·22792	·17053	+·01113	−·09257	−·05127	+·04545	1·05812
·146	·22898	·17061	+·01032	−·09307	−·05078	+·04612	1·05374
·147	·23003	·17069	+·00950	−·09356	−·05028	+·04678	1·04939
·148	·23108	·17076	+·00869	−·09405	−·04978	+·04744	1·04505
·149	·23212	·17082	+·00788	−·09452	−·04928	+·04809	1·04073
·150	·23316	·17087	+·00706	−·09499	−·04877	+·04874	1·03643

TABLE—(*continued*).

$\frac{1}{2}(1-\alpha)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	k
·151	·23419	·17092	+ ·00625	− ·09546	− ·04825	+ ·04938	1·03215
·152	·23522	·17097	+ ·00543	− ·09592	− ·04774	+ ·05002	1·02789
·153	·23625	·17100	+ ·00462	− ·09637	− ·04721	+ ·05065	1·02365
·154	·23727	·17103	+ ·00380	− ·09681	− ·04669	+ ·05127	1·01943
·155	·23829	·17106	+ ·00298	− ·09725	− ·04615	+ ·05189	1·01522
·156	·23930	·17108	+ ·00217	− ·09768	− ·04562	+ ·05250	1·01103
·157	·24031	·17109	+ ·00135	− ·09810	− ·04508	+ ·05311	1·00686
·158	·24131	·17110	+ ·00053	− ·09852	− ·04454	+ ·05371	1·00271
·159	·24232	·17110	− ·00028	− ·09892	− ·04399	+ ·05431	·99858
·160	·24331	·17109	− ·00110	− ·09933	− ·04344	+ ·05490	·99446
·161	·24430	·17108	− ·00191	− ·09972	− ·04288	+ ·05549	·99036
·162	·24529	·17107	− ·00273	− ·10011	− ·04232	+ ·05607	·98627
·163	·24628	·17104	− ·00355	− ·10049	− ·04176	+ ·05664	·98220
·164	·24726	·17102	− ·00436	− ·10087	− ·04120	+ ·05721	·97815
·165	·24823	·17098	− ·00518	− ·10124	− ·04063	+ ·05778	·97411
·166	·24921	·17094	− ·00599	− ·10160	− ·04006	+ ·05834	·97009
·167	·25017	·17090	− ·00681	− ·10196	− ·03948	+ ·05889	·96609
·168	·25114	·17085	− ·00762	− ·10231	− ·03890	+ ·05943	·96210
·169	·25210	·17080	− ·00844	− ·10265	− ·03832	+ ·05998	·95812
·170	·25305	·17073	− ·00925	− ·10299	− ·03774	+ ·06051	·95417
·171	·25401	·17067	− ·01007	− ·10332	− ·03715	+ ·06104	·95022
·172	·25495	·17060	− ·01088	− ·10364	− ·03656	+ ·06156	·94629
·173	·25590	·17052	− ·01169	− ·10396	− ·03597	+ ·06208	·94238
·174	·25684	·17044	− ·01251	− ·10427	− ·03537	+ ·06260	·93848
·175	·25778	·17035	− ·01332	− ·10458	− ·03478	+ ·06310	·93459
·176	·25871	·17026	− ·01413	− ·10487	− ·03417	+ ·06360	·93072
·177	·25964	·17016	− ·01494	− ·10517	− ·03357	+ ·06410	·92686
·178	·26056	·17006	− ·01575	− ·10545	− ·03296	+ ·06459	·92301
·179	·26148	·16995	− ·01656	− ·10573	− ·03236	+ ·06507	·91918
·180	·26240	·16984	− ·01737	− ·10601	− ·03175	+ ·06555	·91537
·181	·26331	·16972	− ·01817	− ·10627	− ·03113	+ ·06603	·91156
·182	·26422	·16960	− ·01898	− ·10653	− ·03052	+ ·06649	·90777
·183	·26513	·16948	− ·01979	− ·10679	− ·02990	+ ·06695	·90399
·184	·26603	·16934	− ·02059	− ·10704	− ·02928	+ ·06741	·90023
·185	·26693	·16921	− ·02140	− ·10728	− ·02866	+ ·06786	·89647
·186	·26782	·16907	− ·02220	− ·10752	− ·02803	+ ·06830	·89273
·187	·26871	·16892	− ·02300	− ·10775	− ·02741	+ ·06874	·88901
·188	·26960	·16877	− ·02380	− ·10798	− ·02678	+ ·06917	·88529
·189	·27049	·16861	− ·02460	− ·10819	− ·02615	+ ·06960	·88159
·190	·27137	·16845	− ·02540	− ·10841	− ·02552	+ ·07002	·87790
·191	·27224	·16829	− ·02620	− ·10861	− ·02489	+ ·07044	·87422
·192	·27311	·16812	− ·02700	− ·10882	− ·02425	+ ·07085	·87055
·193	·27398	·16795	− ·02779	− ·10901	− ·02362	+ ·07125	·86689
·194	·27485	·16777	− ·02859	− ·10920	− ·02298	+ ·07165	·86325
·195	·27571	·16759	− ·02938	− ·10939	− ·02234	+ ·07204	·85962
·196	·27657	·16740	− ·03018	− ·10956	− ·02170	+ ·07243	·85600
·197	·27742	·16721	− ·03097	− ·10974	− ·02106	+ ·07281	·85239
·198	·27827	·16701	− ·03176	− ·10990	− ·02041	+ ·07319	·84879
·199	·27912	·16681	− ·03255	− ·11007	− ·01977	+ ·07356	·84520
·200	·27996	·16661	− ·03334	− ·11022	− ·01912	+ ·07392	·84162

TABLE—(*continued*).

$\frac{1}{2}(1-a)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	h
·201	·28080	·16640	− ·03412	− ·11037	− ·01848	+ ·07428	·83805
·202	·28164	·16619	− ·03491	− ·11051	− ·01783	+ ·07464	·83450
·203	·28247	·16597	− ·03569	− ·11065	− ·01718	+ ·07498	·83095
·204	·28330	·16575	− ·03648	− ·11079	− ·01653	+ ·07532	·82742
·205	·28413	·16553	− ·03726	− ·11091	− ·01587	+ ·07566	·82389
·206	·28495	·16530	− ·03804	− ·11104	− ·01522	+ ·07599	·82038
·207	·28577	·16506	− ·03882	− ·11115	− ·01457	+ ·07632	·81687
·208	·28658	·16483	− ·03959	− ·11126	− ·01391	+ ·07664	·81338
·209	·28739	·16459	− ·04037	− ·11137	− ·01326	+ ·07695	·80990
·210	·28820	·16434	− ·04114	− ·11147	− ·01260	+ ·07726	·80642
·211	·28901	·16409	− ·04192	− ·11157	− ·01194	+ ·07756	·80296
·212	·28981	·16384	− ·04269	− ·11166	− ·01129	+ ·07786	·79950
·213	·29060	·16358	− ·04346	− ·11174	− ·01063	+ ·07815	·79606
·214	·29140	·16332	− ·04423	− ·11182	− ·00997	+ ·07844	·79262
·215	·29219	·16305	− ·04499	− ·11189	− ·00931	+ ·07872	·78919
·216	·29298	·16279	− ·04576	− ·11196	− ·00865	+ ·07899	·78577
·217	·29376	·16251	− ·04652	− ·11203	− ·00799	+ ·07926	·78237
·218	·29454	·16224	− ·04728	− ·11208	− ·00733	+ ·07952	·77897
·219	·29532	·16196	− ·04804	− ·11214	− ·00667	+ ·07978	·77557
·220	·29609	·16167	− ·04880	− ·11218	− ·00600	+ ·08004	·77219
·221	·29686	·16139	− ·04956	− ·11223	− ·00534	+ ·08028	·76882
·222	·29763	·16110	− ·05031	− ·11226	− ·00468	+ ·08052	·76546
·223	·29840	·16080	− ·05107	− ·11230	− ·00402	+ ·08076	·76210
·224	·29916	·16050	− ·05182	− ·11233	− ·00335	+ ·08099	·75875
·225	·29991	·16020	− ·05257	− ·11235	− ·00269	+ ·08122	·75541
·226	·30067	·15990	− ·05332	− ·11237	− ·00203	+ ·08144	·75208
·227	·30142	·15959	− ·05406	− ·11238	− ·00136	+ ·08165	·74876
·228	·30216	·15927	− ·05481	− ·11239	− ·00070	+ ·08186	·74545
·229	·30291	·15896	− ·05555	− ·11239	− ·00004	+ ·08207	·74214
·230	·30365	·15864	− ·05629	− ·11239	+ ·00063	+ ·08226	·73885
·231	·30439	·15832	− ·05703	− ·11238	+ ·00129	+ ·08246	·73556
·232	·30512	·15799	− ·05777	− ·11237	+ ·00195	+ ·08265	·73228
·233	·30585	·15766	− ·05851	− ·11235	+ ·00262	+ ·08283	·72900
·234	·30658	·15733	− ·05924	− ·11233	+ ·00328	+ ·08301	·72574
·235	·30730	·15699	− ·05997	− ·11230	+ ·00394	+ ·08318	·72248
·236	·30802	·15665	− ·06070	− ·11227	+ ·00461	+ ·08334	·71923
·237	·30874	·15631	− ·06143	− ·11224	+ ·00527	+ ·08351	·71599
·238	·30945	·15596	− ·06215	− ·11220	+ ·00593	+ ·08366	·71275
·239	·31017	·15561	− ·06288	− ·11215	+ ·00659	+ ·08381	·70952
·240	·31087	·15526	− ·06360	− ·11210	+ ·00726	+ ·08396	·70630
·241	·31158	·15490	− ·06432	− ·11205	+ ·00792	+ ·08410	·70309
·242	·31228	·15454	− ·06504	− ·11199	+ ·00858	+ ·08423	·69988
·243	·31298	·15418	− ·06576	− ·11192	+ ·00924	+ ·08436	·69668
·244	·31367	·15382	− ·06647	− ·11185	+ ·00990	+ ·08449	·69349
·245	·31436	·15345	− ·06718	− ·11178	+ ·01056	+ ·08461	·69031
·246	·31505	·15308	− ·06789	− ·11170	+ ·01122	+ ·08472	·68713
·247	·31574	·15270	− ·06860	− ·11162	+ ·01188	+ ·08483	·68396
·248	·31642	·15232	− ·06931	− ·11154	+ ·01253	+ ·08494	·68080
·249	·31710	·15194	− ·07001	− ·11145	+ ·01319	+ ·08504	·67764
·250	·31778	·15156	− ·07071	− ·11135	+ ·01385	+ ·08513	·67449

TABLE—(*continued*).

$\frac{1}{2}(1-a)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	h
·251	·31845	·15117	− ·07141	− ·11125	+ ·01450	+ ·08522	·67135
·252	·31912	·15078	− ·07211	− ·11115	+ ·01516	+ ·08530	·66821
·253	·31979	·15039	− ·07280	− ·11104	+ ·01581	+ ·08538	·66508
·254	·32045	·14999	− ·07350	− ·11093	+ ·01647	+ ·08546	·66196
·255	·32111	·14959	− ·07419	− ·11081	+ ·01712	+ ·08553	·65884
·256	·32177	·14919	− ·07488	− ·11069	+ ·01777	+ ·08559	·65573
·257	·32242	·14879	− ·07557	− ·11056	+ ·01842	+ ·08565	·65262
·258	·32307	·14838	− ·07625	− ·11043	+ ·01907	+ ·08571	·64952
·259	·32372	·14797	− ·07693	− ·11030	+ ·01972	+ ·08575	·64643
·260	·32437	·14756	− ·07761	− ·11016	+ ·02037	+ ·08580	·64335
·261	·32501	·14714	− ·07829	− ·11002	+ ·02102	+ ·08584	·64027
·262	·32565	·14672	− ·07897	− ·10987	+ ·02166	+ ·08587	·63719
·263	·32628	·14630	− ·07964	− ·10972	+ ·02231	+ ·08590	·63412
·264	·32691	·14588	− ·08031	− ·10956	+ ·02295	+ ·08593	·63106
·265	·32754	·14545	− ·08098	− ·10940	+ ·02360	+ ·08595	·62801
·266	·32817	·14502	− ·08165	− ·10924	+ ·02424	+ ·08596	·62496
·267	·32879	·14459	− ·08231	− ·10907	+ ·02488	+ ·08597	·62191
·268	·32941	·14415	− ·08298	− ·10890	+ ·02552	+ ·08598	·61887
·269	·33003	·14372	− ·08364	− ·10873	+ ·02616	+ ·08598	·61584
·270	·33065	·14328	− ·08429	− ·10855	+ ·02680	+ ·08598	·61281
·271	·33126	·14283	− ·08495	− ·10837	+ ·02743	+ ·08597	·60979
·272	·33187	·14239	− ·08560	− ·10818	+ ·02807	+ ·08596	·60678
·273	·33247	·14194	− ·08625	− ·10799	+ ·02870	+ ·08594	·60376
·274	·33307	·14149	− ·08690	− ·10779	+ ·02933	+ ·08591	·60076
·275	·33367	·14104	− ·08755	− ·10759	+ ·02997	+ ·08589	·59776
·276	·33427	·14058	− ·08819	− ·10739	+ ·03060	+ ·08586	·59477
·277	·33486	·14012	− ·08883	− ·10718	+ ·03122	+ ·08582	·59178
·278	·33545	·13966	− ·08947	− ·10697	+ ·03185	+ ·08578	·58879
·279	·33604	·13920	− ·09011	− ·10676	+ ·03248	+ ·08573	·58581
·280	·33662	·13873	− ·09074	− ·10654	+ ·03310	+ ·08568	·58284
·281	·33720	·13826	− ·09137	− ·10632	+ ·03372	+ ·08563	·57987
·282	·33778	·13779	− ·09200	− ·10609	+ ·03434	+ ·08557	·57691
·283	·33836	·13732	− ·09263	− ·10587	+ ·03496	+ ·08551	·57395
·284	·33893	·13685	− ·09325	− ·10563	+ ·03558	+ ·08544	·57100
·285	·33950	·13637	− ·09388	− ·10540	+ ·03620	+ ·08536	·56805
·286	·34007	·13589	− ·09450	− ·10516	+ ·03681	+ ·08529	·56511
·287	·34063	·13541	− ·09511	− ·10491	+ ·03743	+ ·08521	·56217
·288	·34119	·13492	− ·09573	− ·10466	+ ·03804	+ ·08512	·55924
·289	·34175	·13443	− ·09634	− ·10441	+ ·03865	+ ·08503	·55631
·290	·34230	·13394	− ·09695	− ·10416	+ ·03926	+ ·08494	·55338
·291	·34286	·13345	− ·09756	− ·10390	+ ·03987	+ ·08484	·55047
·292	·34341	·13296	− ·09816	− ·10364	+ ·04047	+ ·08473	·54755
·293	·34395	·13246	− ·09876	− ·10337	+ ·04107	+ ·08463	·54464
·294	·34449	·13196	− ·09936	− ·10310	+ ·04168	+ ·08451	·54174
·295	·34503	·13146	− ·09996	− ·10283	+ ·04228	+ ·08440	·53884
·296	·34557	·13096	− ·10056	− ·10256	+ ·04287	+ ·08428	·53594
·297	·34611	·13046	− ·10115	− ·10228	+ ·04347	+ ·08415	·53305
·298	·34664	·12995	− ·10174	− ·10199	+ ·04407	+ ·08402	·53016
·299	·34717	·12944	− ·10233	− ·10171	+ ·04466	+ ·08389	·52728
·300	·34769	·12893	− ·10291	− ·10142	+ ·04525	+ ·08375	·52440

TABLE—(*continued*).

$\frac{1}{2}(1-\alpha)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	h
·301	·34822	·12841	− ·10349	− ·10113	+ ·04584	+ ·08361	·52153
·302	·34874	·12790	− ·10407	− ·10083	+ ·04643	+ ·08347	·51866
·303	·34925	·12738	− ·10465	− ·10053	+ ·04701	+ ·08332	·51579
·304	·34977	·12686	− ·10522	− ·10023	+ ·04759	+ ·08316	·51293
·305	·35028	·12634	− ·10580	− ·09992	+ ·04818	+ ·08301	·51007
·306	·35079	·12581	− ·10636	− ·09961	+ ·04876	+ ·08284	·50722
·307	·35129	·12529	− ·10693	− ·09930	+ ·04933	+ ·08268	·50437
·308	·35180	·12476	− ·10750	− ·09899	+ ·04991	+ ·08251	·50153
·309	·35230	·12423	− ·10806	− ·09867	+ ·05048	+ ·08233	·49869
·310	·35279	·12370	− ·10862	− ·09834	+ ·05105	+ ·08216	·49585
·311	·35329	·12316	− ·10917	− ·09802	+ ·05162	+ ·08197	·49302
·312	·35378	·12263	− ·10973	− ·09769	+ ·05219	+ ·08179	·49019
·313	·35427	·12209	− ·11028	− ·09736	+ ·05276	+ ·08160	·48736
·314	·35475	·12155	− ·11082	− ·09703	+ ·05332	+ ·08140	·48454
·315	·35524	·12101	− ·11137	− ·09669	+ ·05388	+ ·08121	·48173
·316	·35572	·12046	− ·11191	− ·09635	+ ·05444	+ ·08101	·47891
·317	·35620	·11992	− ·11245	− ·09600	+ ·05500	+ ·08080	·47610
·318	·35667	·11937	− ·11299	− ·09566	+ ·05555	+ ·08059	·47330
·319	·35714	·11882	− ·11353	− ·09531	+ ·05610	+ ·08038	·47050
·320	·35761	·11827	− ·11406	− ·09495	+ ·05665	+ ·08016	·46770
·321	·35808	·11771	− ·11459	− ·09460	+ ·05720	+ ·07994	·46490
·322	·35854	·11716	− ·11512	− ·09424	+ ·05775	+ ·07972	·46211
·323	·35900	·11660	− ·11564	− ·09388	+ ·05829	+ ·07949	·45933
·324	·35946	·11604	− ·11616	− ·09351	+ ·05883	+ ·07926	·45654
·325	·35991	·11548	− ·11668	− ·09315	+ ·05937	+ ·07902	·45376
·326	·36037	·11492	− ·11720	− ·09278	+ ·05991	+ ·07878	·45099
·327	·36082	·11436	− ·11771	− ·09240	+ ·06044	+ ·07854	·44821
·328	·36126	·11379	− ·11822	− ·09203	+ ·06097	+ ·07829	·44544
·329	·36171	·11322	− ·11873	− ·09165	+ ·06150	+ ·07804	·44268
·330	·36215	·11265	− ·11923	− ·09127	+ ·06203	+ ·07779	·43991
·331	·36259	·11208	− ·11974	− ·09088	+ ·06255	+ ·07753	·43715
·332	·36302	·11151	− ·12024	− ·09049	+ ·06308	+ ·07727	·43440
·333	·36346	·11093	− ·12073	− ·09010	+ ·06360	+ ·07701	·43164
·334	·36389	·11036	− ·12123	− ·08971	+ ·06412	+ ·07674	·42889
·335	·36431	·10978	− ·12172	− ·08932	+ ·06463	+ ·07647	·42615
·336	·36474	·10920	− ·12221	− ·08892	+ ·06514	+ ·07620	·42340
·337	·36516	·10862	− ·12270	− ·08852	+ ·06565	+ ·07592	·42066
·338	·36558	·10804	− ·12318	− ·08811	+ ·06616	+ ·07564	·41793
·339	·36600	·10745	− ·12366	− ·08771	+ ·06667	+ ·07535	·41519
·340	·36641	·10687	− ·12414	− ·08730	+ ·06717	+ ·07507	·41246
·341	·36682	·10628	− ·12461	− ·08689	+ ·06767	+ ·07477	·40974
·342	·36723	·10569	− ·12509	− ·08647	+ ·06817	+ ·07448	·40701
·343	·36764	·10510	− ·12555	− ·08606	+ ·06867	+ ·07418	·40429
·344	·36804	·10451	− ·12602	− ·08564	+ ·06916	+ ·07388	·40157
·345	·36844	·10391	− ·12649	− ·08522	+ ·06965	+ ·07358	·39886
·346	·36884	·10332	− ·12695	− ·08479	+ ·07014	+ ·07327	·39614
·347	·36923	·10272	− ·12741	− ·08437	+ ·07062	+ ·07296	·39343
·348	·36962	·10212	− ·12786	− ·08394	+ ·07110	+ ·07264	·39073
·349	·37001	·10152	− ·12831	− ·08351	+ ·07158	+ ·07232	·38802
·350	·37040	·10092	− ·12876	− ·08307	+ ·07206	+ ·07200	·38532

TABLE—(*continued*).

$\frac{1}{2}(1-a)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	h
·351	·37078	·10032	− ·12921	− ·08264	+ ·07254	+ ·07168	·38262
·352	·37116	·09971	− ·12966	− ·08220	+ ·07301	+ ·07135	·37993
·353	·37154	·09911	− ·13010	− ·08176	+ ·07348	+ ·07102	·37723
·354	·37192	·09850	− ·13054	− ·08131	+ ·07395	+ ·07069	·37454
·355	·37229	·09789	− ·13097	− ·08087	+ ·07441	+ ·07035	·37186
·356	·37266	·09728	− ·13140	− ·08042	+ ·07487	+ ·07002	·36917
·357	·37303	·09667	− ·13183	− ·07997	+ ·07533	+ ·06967	·36649
·358	·37340	·09606	− ·13226	− ·07952	+ ·07579	+ ·06933	·36381
·359	·37376	·09544	− ·13269	− ·07906	+ ·07624	+ ·06898	·36113
·360	·37412	·09483	− ·13311	− ·07861	+ ·07669	+ ·06863	·35846
·361	·37447	·09421	− ·13353	− ·07815	+ ·07714	+ ·06827	·35579
·362	·37483	·09359	− ·13394	− ·07768	+ ·07758	+ ·06792	·35312
·363	·37518	·09297	− ·13436	− ·07722	+ ·07803	+ ·06756	·35045
·364	·37553	·09235	− ·13477	− ·07675	+ ·07847	+ ·06719	·34779
·365	·37588	·09173	− ·13517	− ·07629	+ ·07890	+ ·06683	·34513
·366	·37622	·09111	− ·13558	− ·07582	+ ·07934	+ ·06646	·34247
·367	·37656	·09048	− ·13598	− ·07534	+ ·07977	+ ·06609	·33981
·368	·37690	·08985	− ·13638	− ·07487	+ ·08020	+ ·06571	·33715
·369	·37724	·08923	− ·13677	− ·07439	+ ·08062	+ ·06534	·33450
·370	·37757	·08860	− ·13717	− ·07391	+ ·08105	+ ·06496	·33185
·371	·37790	·08797	− ·13756	− ·07343	+ ·08147	+ ·06458	·32921
·372	·37823	·08734	− ·13794	− ·07295	+ ·08188	+ ·06419	·32656
·373	·37855	·08671	− ·13833	− ·07246	+ ·08230	+ ·06380	·32392
·374	·37888	·08607	− ·13871	− ·07198	+ ·08271	+ ·06341	·32128
·375	·37920	·08544	− ·13909	− ·07149	+ ·08312	+ ·06302	·31864
·376	·37951	·08480	− ·13946	− ·07100	+ ·08352	+ ·06262	·31600
·377	·37983	·08416	− ·13984	− ·07050	+ ·08392	+ ·06222	·31337
·378	·38014	·08353	− ·14021	− ·07001	+ ·08432	+ ·06182	·31074
·379	·38045	·08289	− ·14057	− ·06951	+ ·08472	+ ·06142	·30811
·380	·38076	·08225	− ·14094	− ·06901	+ ·08512	+ ·06101	·30548
·381	·38106	·08160	− ·14130	− ·06851	+ ·08551	+ ·06061	·30286
·382	·38136	·08096	− ·14166	− ·06801	+ ·08589	+ ·06019	·30023
·383	·38166	·08032	− ·14201	− ·06750	+ ·08628	+ ·05978	·29761
·384	·38196	·07967	− ·14236	− ·06700	+ ·08666	+ ·05936	·29499
·385	·38225	·07903	− ·14271	− ·06649	+ ·08704	+ ·05895	·29237
·386	·38254	·07838	− ·14306	− ·06598	+ ·08742	+ ·05853	·28976
·387	·38283	·07773	− ·14340	− ·06547	+ ·08779	+ ·05810	·28715
·388	·38312	·07708	− ·14374	− ·06495	+ ·08816	+ ·05766	·28454
·389	·38340	·07643	− ·14408	− ·06444	+ ·08853	+ ·05725	·28193
·390	·38368	·07578	− ·14442	− ·06392	+ ·08889	+ ·05682	·27932
·391	·38396	·07513	− ·14475	− ·06340	+ ·08925	+ ·05638	·27671
·392	·38423	·07447	− ·14508	− ·06288	+ ·08961	+ ·05595	·27411
·393	·38451	·07382	− ·14540	− ·06236	+ ·08997	+ ·05551	·27151
·394	·38478	·07316	− ·14573	− ·06183	+ ·09032	+ ·05507	·26891
·395	·38504	·07251	− ·14604	− ·06131	+ ·09067	+ ·05463	·26631
·396	·38531	·07185	− ·14636	− ·06078	+ ·09101	+ ·05419	·26371
·397	·38557	·07119	− ·14668	− ·06025	+ ·09136	+ ·05374	·26112
·398	·38583	·07053	− ·14699	− ·05972	+ ·09170	+ ·05329	·25853
·399	·38609	·06987	− ·14730	− ·05919	+ ·09203	+ ·05284	·25594
·400	·38634	·06921	− ·14760	− ·05866	+ ·09237	+ ·05239	·25335

TABLE—(*continued*).

$\frac{1}{2}(1-\alpha)$	r_1	r_2	r_3	r_4	r_5	r_6	h
·401	·38659	·06855	− ·14790	− ·05812	+ ·09270	+ ·05193	·25076
·402	·38684	·06789	− ·14820	− ·05758	+ ·09303	+ ·05148	·24817
·403	·38709	·06722	− ·14850	− ·05705	+ ·09335	+ ·05102	·24559
·404	·38734	·06656	− ·14879	− ·05651	+ ·09367	+ ·05056	·24301
·405	·38758	·06589	− ·14908	− ·05596	+ ·09399	+ ·05010	·24043
·406	·38782	·06522	− ·14937	− ·05542	+ ·09430	+ ·04963	·23785
·407	·38805	·06456	− ·14965	− ·05488	+ ·09462	+ ·04916	·23527
·408	·38829	·06389	− ·14993	− ·05433	+ ·09493	+ ·04869	·23269
·409	·38852	·06322	− ·15021	− ·05378	+ ·09523	+ ·04822	·23012
·410	·38875	·06255	− ·15049	− ·05323	+ ·09553	+ ·04775	·22754
·411	·38897	·06188	− ·15076	− ·05268	+ ·09583	+ ·04728	·22497
·412	·38920	·06121	− ·15103	− ·05213	+ ·09613	+ ·04680	·22240
·413	·38942	·06053	− ·15130	− ·05158	+ ·09642	+ ·04632	·21983
·414	·38964	·05986	− ·15156	− ·05102	+ ·09671	+ ·04584	·21727
·415	·38985	·05919	− ·15182	− ·05047	+ ·09700	+ ·04536	·21470
·416	·39007	·05851	− ·15208	− ·04991	+ ·09728	+ ·04488	·21214
·417	·39028	·05784	− ·15233	− ·04935	+ ·09756	+ ·04439	·20957
·418	·39049	·05716	− ·15258	− ·04879	+ ·09784	+ ·04390	·20701
·419	·39069	·05648	− ·15283	− ·04823	+ ·09811	+ ·04341	·20445
·420	·39089	·05580	− ·15308	− ·04767	+ ·09838	+ ·04292	·20189
·421	·39109	·05513	− ·15332	− ·04711	+ ·09865	+ ·04243	·19934
·422	·39129	·05445	− ·15356	− ·04654	+ ·09891	+ ·04194	·19678
·423	·39149	·05377	− ·15380	− ·04598	+ ·09918	+ ·04144	·19422
·424	·39168	·05309	− ·15403	− ·04541	+ ·09943	+ ·04094	·19167
·425	·39187	·05240	− ·15426	− ·04484	+ ·09969	+ ·04044	·18912
·426	·39206	·05172	− ·15449	− ·04427	+ ·09994	+ ·03994	·18657
·427	·39224	·05104	− ·15471	− ·04370	+ ·10019	+ ·03944	·18402
·428	·39243	·05036	− ·15493	− ·04313	+ ·10043	+ ·03894	·18147
·429	·39261	·04967	− ·15515	− ·04256	+ ·10067	+ ·03843	·17892
·430	·39279	·04899	− ·15537	− ·04198	+ ·10091	+ ·03793	·17637
·431	·39296	·04830	− ·15558	− ·04141	+ ·10115	+ ·03742	·17383
·432	·39313	·04761	− ·15579	− ·04083	+ ·10138	+ ·03691	·17128
·433	·39330	·04693	− ·15599	− ·04026	+ ·10161	+ ·03640	·16874
·434	·39347	·04624	− ·15620	− ·03968	+ ·10183	+ ·03589	·16620
·435	·39364	·04555	− ·15640	− ·03910	+ ·10205	+ ·03537	·16366
·436	·39380	·04486	− ·15659	− ·03852	+ ·10227	+ ·03486	·16112
·437	·39396	·04418	− ·15679	− ·03794	+ ·10249	+ ·03434	·15858
·438	·39411	·04349	− ·15698	− ·03735	+ ·10270	+ ·03382	·15604
·439	·39427	·04280	− ·15717	− ·03677	+ ·10291	+ ·03330	·15351
·440	·39442	·04211	− ·15735	− ·03619	+ ·10311	+ ·03278	·15097
·441	·39457	·04141	− ·15753	− ·03560	+ ·10331	+ ·03226	·14843
·442	·39472	·04072	− ·15771	− ·03502	+ ·10351	+ ·03174	·14590
·443	·39486	·04003	− ·15789	− ·03443	+ ·10371	+ ·03121	·14337
·444	·39501	·03934	− ·15806	− ·03384	+ ·10390	+ ·03069	·14084
·445	·39514	·03864	− ·15823	− ·03325	+ ·10409	+ ·03016	·13830
·446	·39528	·03795	− ·15840	− ·03266	+ ·10427	+ ·02963	·13577
·447	·39542	·03726	− ·15856	− ·03207	+ ·10446	+ ·02910	·13324
·448	·39555	·03656	− ·15872	− ·03148	+ ·10463	+ ·02858	·13072
·449	·39568	·03587	− ·15888	− ·03089	+ ·10481	+ ·02804	·12819
·450	·39580	·03517	− ·15904	− ·03030	+ ·10498	+ ·02751	·12566

TABLE—(*continued*).

$\frac{1}{2}(1-a)$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	h
·451	·39593	·03447	−·15919	−·02970	+·10515	+·02698	·12314
·452	·39605	·03378	−·15934	−·02911	+·10532	+·02644	·12061
·453	·39617	·03308	−·15948	−·02851	+·10548	+·02591	·11809
·454	·39629	·03238	−·15962	−·02792	+·10564	+·02537	·11556
·455	·39640	·03168	−·15976	−·02732	+·10579	+·02484	·11304
·456	·39651	·03099	−·15990	−·02673	+·10594	+·02430	·11052
·457	·39662	·03029	−·16003	−·02613	+·10609	+·02376	·10799
·458	·39673	·02959	−·16016	−·02553	+·10624	+·02322	·10547
·459	·39683	·02889	−·16029	−·02493	+·10638	+·02268	·10295
·460	·39694	·02819	−·16041	−·02433	+·10652	+·02214	·10043
·461	·39703	·02749	−·16053	−·02373	+·10665	+·02159	·09791
·462	·39713	·02679	−·16065	−·02313	+·10678	+·02105	·09540
·463	·39723	·02609	−·16077	−·02253	+·10691	+·02051	·09288
·464	·39732	·02539	−·16088	−·02193	+·10704	+·01996	·09036
·465	·39741	·02469	−·16099	−·02132	+·10716	+·01941	·08784
·466	·39749	·02398	−·16109	−·02072	+·10727	+·01887	·08533
·467	·39758	·02328	−·16120	−·02012	+·10739	+·01832	·08281
·468	·39766	·02258	−·16130	−·01951	+·10750	+·01777	·08030
·469	·39774	·02188	−·16139	−·01891	+·10761	+·01722	·07778
·470	·39781	·02117	−·16149	−·01830	+·10771	+·01668	·07527
·471	·39789	·02047	−·16158	−·01770	+·10781	+·01613	·07276
·472	·39796	·01977	−·16166	−·01709	+·10791	+·01558	·07024
·473	·39803	·01906	−·16175	−·01648	+·10801	+·01502	·06773
·474	·39809	·01836	−·16183	−·01588	+·10810	+·01447	·06522
·475	·39816	·01765	−·16191	−·01527	+·10818	+·01392	·06271
·476	·39822	·01695	−·16198	−·01466	+·10827	+·01337	·06020
·477	·39828	·01625	−·16206	−·01405	+·10835	+·01281	·05768
·478	·39834	·01554	−·16212	−·01344	+·10842	+·01226	·05517
·479	·39839	·01484	−·16219	−·01284	+·10850	+·01171	·05266
·480	·39844	·01413	−·16225	−·01223	+·10857	+·01115	·05015
·481	·39849	·01342	−·16231	−·01162	+·10864	+·01060	·04764
·482	·39854	·01272	−·16237	−·01101	+·10870	+·01004	·04513
·483	·39858	·01201	−·16242	−·01040	+·10876	+·00949	·04263
·484	·39862	·01131	−·16247	−·00979	+·10882	+·00893	·04012
·485	·39866	·01060	−·16252	−·00918	+·10887	+·00837	·03761
·486	·39870	·00990	−·16257	−·00857	+·10892	+·00782	·03510
·487	·39873	·00919	−·16261	−·00796	+·10896	+·00726	·03259
·488	·39876	·00848	−·16265	−·00734	+·10901	+·00670	·03008
·489	·39879	·00778	−·16268	−·00673	+·10905	+·00614	·02758
·490	·39882	·00707	−·16271	−·00612	+·10908	+·00559	·02507
·491	·39884	·00636	−·16274	−·00551	+·10912	+·00503	·02256
·492	·39886	·00566	−·16277	−·00490	+·10914	+·00447	·02005
·493	·39888	·00495	−·16279	−·00429	+·10917	+·00391	·01755
·494	·39890	·00424	−·16281	−·00367	+·10919	+·00335	·01504
·495	·39891	·00354	−·16283	−·00306	+·10921	+·00279	·01253
·496	·39892	·00283	−·16284	−·00245	+·10923	+·00224	·01003
·497	·39893	·00212	−·16285	−·00184	+·10924	+·00168	·00752
·498	·39894	·00141	−·16286	−·00122	+·10925	+·00112	·00501
·499	·39894	·00071	−·16287	−·00061	+·10925	+·00056	·00251
·500	·39894	·00000	−·16287	·00000	+·10925	·00000	·00000

ON THE SELECTIVE ELIMINATION OCCURRING DURING THE DEVELOPMENT OF THE FRUITS OF STAPHYLEA.

By J. ARTHUR HARRIS, Ph.D., Carnegie Institution of Washington.

CONTENTS.

		Page
I.	Introduction	453
II.	Material	454
III.	Statement of Problems and Methods of Investigation	458
IV.	Discussion of Problems on the Basis of Actual Data for 1908	459
	Problem 1. Is there a difference in the mean number of ovules per locule, or per fruit, in the ovaries which are eliminated and those which develop to maturity?	459
	Problem 2. Is there a difference in the variability in number of ovules per locule and per fruit in the eliminated ovaries and those which develop into mature fruits?	468
	Standard Deviations	470
	Coefficients of Variation	473
	Problem 3. Is there a difference in the radial asymmetry (as indicated by the number of ovules per locule) in the eliminated ovaries and those which develop to maturity?	476
	Problem 4. Is there a difference in the relative number of locules with "odd" numbers of ovules in the eliminated ovaries and those which develop to maturity?	483
	Problem 5. Is there a difference in the percentage of 2- and 4-merous ovaries among the eliminated ovaries and those which develop to maturity?	486
V.	Discussion of Problems on the Basis of the Actual Data for 1906	487
VI.	Interrelationship of Characters Considered	491
	1. The correlation between the coefficient of asymmetry and the number of "odd" locules per ovary	491
	2. The correlation between the coefficient of asymmetry and the number of ovules per locule	493
	3. The correlation between the number of "odd" locules per ovary and the number of ovules per locule	494
VII.	Recapitulation	494
VIII.	Appendix of Tables of Data	495

I. INTRODUCTION.

Volumes have been written concerning Natural Selection—its existence or non-existence, its potency or its inefficiency—but much to the discredit of biologists, thousands of pages of this literature may be explored without finding a single oasis of quantitative data to relieve the monotony of the polemic desert.

Fortunately there are a few notable exceptions, among which memoirs by Weldon*†, Bumpus‡, Beeton and Pearson§, Weldon‖, Di Cesnola¶, Crampton**, Di Cesnola††, Kellogg and Bell‡‡ and perhaps a few others may be mentioned as contributing to the direct quantitative evidence concerning the potency of natural selection.

These researches, highly valuable in themselves, are only a slight beginning in the scientific investigation of a field of cardinal biological importance. Perhaps in the earlier days of Darwinism it was not possible to do much more than collect qualitative evidence and to reason from analogy; one great value of the researches just cited is that they show that the exceedingly complex problems of natural selection can be made the field of quantitative investigation, and now that this fact has been demonstrated I think that we should stigmatize as merely pseudo-scientific discussions of selection not based on quantitative data.

Some years ago it occurred to me that it might be possible to ascertain some facts concerning selective elimination by comparing the characters of ovaries which develop into mature fruits with those which do not. That this is a distinct problem from that of the selective elimination of individuals I am quite aware. It is a consideration of the fitness of certain (morphological) types of organs

* Weldon, W. F. R.: "An attempt to measure the death-rate due to the selective elimination of *Carcinus moenas* with respect to a particular dimension." *Proc. Roy. Soc. Lond.* Vol. LVII. pp. 360—379, 1895.

† Weldon, W. F. R.: Presidential address, Zoological Section. *Trans. Brit. Assoc. Bristol*, 1898, pp. 887—902.

‡ Bumpus, H. C.: "The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*." *Biol. Lect. Del. Mar. Biol. Lab. Wood's Holl*, 1897—98, pp. 209—226, 1899.

§ Beeton, Mary, and K. Pearson: "Data for the problem of evolution in man. II. A first study of the inheritance of longevity, and the selective death-rate in man." *Proc. Roy. Soc.* Vol. LXV. pp. 290—305, 1899.

‖ Weldon, W. F. R.: "A first study of natural selection in *Clausilia laminata*." *Biometrika*, Vol. I. pp. 109—124, 1901.

¶ Di Cesnola, A. P.: "Preliminary note on the protective value of colour in *Mantis religiosa*." *Biometrika*, Vol. III. pp. 58, 59, 1903.

** Crampton, H. E.: "Experimental and statistical studies upon Lepidoptera. I. Variation and elimination in *Philosamia cynthia*." *Biometrika*, Vol. III. pp. 113—130, 1904. See also *Journ. Exp. Zool.* Vol. II. pp. 425—430, 1905.

†† Di Cesnola, A. P.: "A first study of natural selection in *Helix arbustorum*." *Biometrika*, Vol. V. pp. 387—399, 1907.

‡‡ Kellogg, V. L. and Ruby G. Bell: "Studies of variation in insects." *Proc. Wash. Acad. Sci.* Vol. VI. pp. 324—329, 1905.

produced by individuals rather than of the fitness of the individuals with given peculiarities. All will admit, I believe, that from the morphological and physiological view-point this problem is well worthy of careful study. In comparison with the results which may be obtained in quantitative investigation of selective elimination of individuals, the results should have real interest to students of evolution.

A first study was made of the ovaries of the leguminous plant *Cercis**. Here the results were negative, but there were certain limitations to the material which I have never succeeded in removing. The results were published for two reasons. First, I wanted if possible to interest others in a problem which needs the results of several independent researches for its solution. Second, I had the present investigation nearly far enough along to see what the outcome would be, and I wanted to draw the conclusions from my first set of data quite independently and entirely unbiased by the evidence of a second study.

II. Material.

The fruit of the American Bladder Nut, *Staphylea trifolia*, is familiar to botanists as an inflated, membranaceous capsule of (generally) three locules, the dorsal sutures of which project to form three lobes. The carpels are (generally) united throughout the greater portion of the length of the fruit. From 4 to

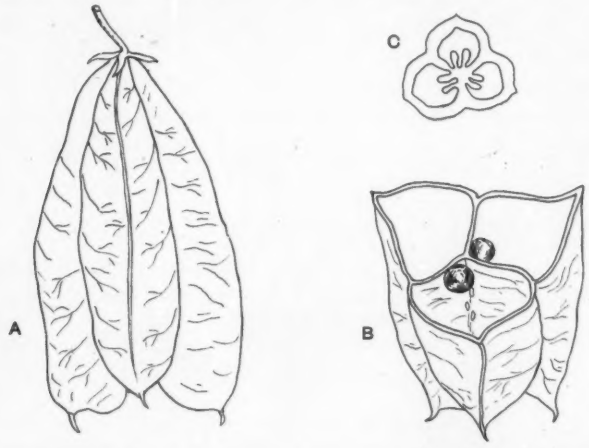


Fig. 1. Diagram showing the external appearance (A) and cross section (B) of fruit of *Staphylea*. Note the one seed and the arrangement of the aborted ovules in the front locule. The cross section of a young ovary (C) shows diagrammatically the disposition of the ovules of each locule in two rows along the inner angle.

* Harris, J. Arthur: "Is there a Selective Elimination in the Fruiting of the Leguminosae?" *Am. Nat.* Vol. xliii. pp. 556—559, 1909.

12 ovules are borne in two series along the inner angle of the locule. None, one or a few of the ovules in each cell or carpel develop into smooth, bony seeds.

The flowers are borne in racemes or panicles. Only a small percentage of those formed in an inflorescence mature fruits. I am able to present no reliable statistics on this point, but I would say that in the individuals which I have studied not over 10 per cent. of the flowers of an inflorescence which ripens fruits at all mature their ovaries. What proportion of inflorescences fail entirely to ripen fruits I am quite unable to say.

It has not seemed worth while to secure quantitative data on this point. The exact proportion varies from season to season. It is always so small that a very stringent elimination is obvious.

This memoir is limited to two series of data, (1) that especially collected for the purpose in 1908, and (2) a small lot of ovaries in different stages of development taken for other purposes in the spring of 1906.

1. *The* 1908 *Collections.*

The collection of this material and the counting of the number of ovules per locule was carried out by my two assistants, Miss Eva N. Dixon and Miss Rose M. Pechmann, and my sister, Miss Nellie L. Harris. I regret not having been able to attend to the sampling personally, but the necessity for an investigation of the problem of selective elimination in *Staphylea* became apparent to me after I had left the Missouri Botanical Garden, and the necessity of devoting the spring season to experimental plantings at the Station for Experimental Evolution, precluded my visiting the Garden at the flowering season for *Staphylea* to make the collections myself.

As for the countings, I cannot state too highly my indebtedness to these assistants for the faithful way in which the arduous task was carried out. In all over 21,000 locules had to be opened and the number of ovules counted. For the mature fruits the opening was simple, but the counting of the aborted ovules is not always an easy task. For the immature ovaries the work was much more trying to the patience and eyesight. It was quite out of the question to deal with the fresh material, and the difficulty of dissection and counting was increased by preservation in alcohol. First the ovaries had to be dissected out of the flowers and then each of the locules opened with a needle and the number of ovules counted. Possibly some of my fellow botanists would have had the deftness of fingers and the patience to carry through the countings, but I would have found the task impossible, and even Miss Pechmann and Miss Dixon occasionally nearly lost heart during the several months the work was under way.

The collection of material to determine whether or no there is a selective elimination is a difficult problem in a form like *Staphylea*. The number of flowers produced is very large, and it is quite impossible to take account of

even a small proportion of them. The inflorescence is long enough to allow the lower ovaries to be developing into young fruits while the upper ones are still in the unopened buds. A collection of the flowers before any elimination has taken place cannot be made for comparison with the ripened fruits, for before the more distal flowers are mature enough for the ovules to be counted with safety, the more proximal ones may be falling from the plant or developing into young fruits.

It is desirable that the sample represent not only all parts of the plant, but also *all regions of the inflorescence*, since, should the ovaries be correlated with their position on the axis, the constants from material carelessly collected might not at all represent the population from which it is drawn.

The labour of gathering ovaries with a record of their position on the inflorescence seemed to be prohibitive. After careful consideration it became apparent that the collection of the spring material must be carried out in a rather rough and ready manner. The following was adopted.

When a shrub was well in flower it was shaken gently to dislodge the flowers which had ceased to develop and were ready to fall. Sometimes flowers were on the ground which had been dislodged by wind and rain, but these were never taken for fear that they had been removed by too great violence. After the tree had been shaken and the fallen sample collected, a gathering as nearly random as possible of the opened flowers which remained was taken.

In a few cases only a single collection of fallen and apparently developing flowers was made from an individual, but usually the shrub was visited again after a few days and a second gathering was made in the same manner as the first.

Finally a series of matured fruits from these individuals was taken in the autumn.

In the two spring series only ovaries dissected from well-opened flowers were examined. All buds were discarded for fear that the ovules might not yet all be laid down. By discarding any young fruits which fell we avoided one danger of overweighting the proximal region of the inflorescence.

From the foregoing remarks it appears that the 1908 ovaries fall into three classes:

A. Ovaries from opened flowers which were eliminated from the shrub.

B. Ovaries from opened flowers which were apparently continuing their development at the time of sampling.

C. Ovaries which had completed their development to mature fruits.

Series B should represent the original population of ovaries more nearly than A, since it contains many which would be eliminated later. Because of the methods necessarily employed in collecting the samples, I do not look upon the comparison of Series A and Series B as very critical. Only a relatively small

proportion of the ovaries develop to maturity; Series A doubtless contains many which are non-selectively eliminated, while B includes many which a little later would have fallen from the tree and been classified as A.

It seems worth while, however, to carry out the comparison between Series A and B. The critical tests are those between B and C and A and C, the most significant being the latter.

In *Staphylea* the normal type of ovary is 3-merous. Occasionally individuals are found which produce a large proportion of 4-merous fruits. Such a one is shrub 36, which was omitted in all these comparisons because I wanted to deal with as nearly a "normal" population of 3-merous ovaries as possible. I feared that the 3-merous ovaries produced on a shrub which bore a very high percentage of 4-merous fruits might be in some characteristics—morphological or physiological—different from those formed by the growing points of an individual with almost exclusively trimerous tendencies.

Shrub 28 bore no flower in 1908.

In dealing with the raw data it has seemed desirable to treat the material for each individual separately, and for the following reason. Rather extensive experience in dealing with biometric constants calculated from the fruits of individuals has shown that there may be very real differences between those of different plants. If each individual did not contribute the same quantity of material to the three collections studied, it is quite possible that in dealing with lumped samples differences in constants might arise quite independently of any influence of selective elimination. I think that with 28 individuals the danger of error from this source is not great, but conclusions based on the results from 28 individual small samples will be stronger than those resting upon merely one large composite series.

2. *The* 1906 *Collections.*

As indicated above these collections were not made primarily for a study of selective elimination at all, but for a study of some questions of fertility and fecundity. Constants calculated from them indicated the necessity for the detailed study attempted in 1908. The results are treated here as supplementary to those from the more detailed investigations made later.

The collection comprises 270 inflorescences* bearing partly developed fruits, none of which had reached a length greater than 20 mm., and was made primarily to determine the change in the correlation between the number of ovules formed and the number developing into seeds at different stages in the development of the fruit. The fruit increases in size very rapidly during its early stages of development, and only a rough measure of its size was considered worth while. Three classes, 6—10 mm., 11—15 mm. and 16—20 mm. in length, were recognized†. The

* The correlation between the number of flowers formed and the number of fruits still continuing to develop at the time the inflorescences were examined is discussed in *Biometrika*, Vol. VI. pp. 440, 441, 1909.

† These are exclusive of the short styles.

shape of the fruit varies considerably, and the length affords only a rough measure of the size, but it is ample for present purposes.

Fruits which have reached a length of 11—15 mm. may be considered to be a selected class as compared with those 6—10 mm. in length—if selection occurs at all—and those of 16—20 mm. in length may be regarded as a selected class as compared with both of the lower classes.

III. Statement of Problems and Methods of Investigation.

In this memoir five characters of the ovary are considered. They are:

(1) Number of ovules per locule.

(2) Total number of ovules per ovary.

(3) Radial asymmetry of the ovary with respect to the number of ovules per locule.

(4) Number of locules per ovary with an "odd" number of ovules.

(5) Number of locules per ovary.

To demonstrate that there is a selective elimination for any or all of these characters, it is necessary to show that there are differences between the constants for the ovaries which develop into fruits and those which do not, of such an order of magnitude that they cannot reasonably be referred to the probable errors of random sampling. In these pages I follow the rather common example of statisticians in regarding differences of at least 2·5 times their probable errors as significant. In addition I have demanded reasonably constant results from the series of individual plants.

The working hypothesis of the existence of a selective elimination for any character whatever may be tested by determining the difference between the constants for:

(*a*) A sample from an original population before elimination and a sample of the eliminated individuals.

(*b*) A sample from an original population before elimination and a sample of the individuals remaining after elimination.

(*c*) A sample of eliminated individuals and a sample of individuals unaffected by the elimination.

Here "individual," "sample" and "population" are used in their commonly accepted and convenient statistical sense. By individual we understand a particular ovary, by sample the collection of ovaries with which the limitations of time permitted us to deal, and by population the total ovaries produced by a given shrub, or the total produced by all the shrubs investigated, as the case may be.

A constant calculated from a sample does not necessarily describe precisely the population from which it was drawn, for it may be too large or too small by an amount depending upon the chance errors of random sampling. As biologists we are interested in our samples only in so far as they indicate for us the characteristics of the population. The degree of significance to be attached to the constant calculated from a sample as a description of a population as a whole is determined by the probable error of the constant. The probable error of random sampling has nothing whatever to do with the care of the naturalist in doing his work, as some biologists seem to believe. Be the observer never so keen and conscientious, the significance to be attached to his conclusions is always to some extent dependent upon the number of observations before him on his work-table. This is merely a conclusion drawn by common sense from common observation. Fortunately in many fields of work the mathematician can give us formulae for determining the relative weight to be attached to constants calculated from different masses of data. These words of explanation are inserted for the benefit of fellow-biologists who may still be unacquainted with the true nature of the probable error of random sampling so much used by biometricians, for throughout these pages I find it necessary to return continually to this fundamental safeguard in the interpretation of the significance, as descriptions of populations, of constants drawn from samples of the population.

In the investigation of the problem of selective elimination on the *Cercis* material, I was necessarily limited to the comparison of the constants for the original population with those for the eliminated ovaries, since a severe frost prevented any of the ovaries maturing. Here we are able to compare eliminated ovaries, the original population after partial elimination, and the matured fruits.

Further details concerning problems and methods will be taken up under the individual problems, to which we now pass.

IV. Discussion of Problems on Basis of Actual Data for 1908.

Problem 1. Is there a difference in the mean number of ovules per locule, or per fruit, in the ovaries which are eliminated and those which develop to maturity?

Seriations of the number of ovules per locule are given for these collections of material in Appendix, Tables XXVIII—XXX. The data for the total ovules per fruit are given for the three collections from the 28 individuals in Tables XXXI—XXXIII. The means, standard deviations and coefficients of variation for the number of ovules per locule, and the standard deviations and coefficients of variation of the total ovules per ovary are given for the 28 individuals of the three collections and their totals in Tables I—III.

In comparing the means for the three series it is unnecessary to treat both ovules per locule and total ovules per fruit, for the latter is necessarily three times

TABLE I.

Physical Constants for Series A.

Number of Shrub	Ovules per Locule			Ovules per Ovary	
	Average and Probable Error	Standard Deviation and Probable Error	Coefficient of Variation	Standard Deviation and Probable Error	Coefficient of Variation
11	8·117 ± ·022	·574 ± ·016	7·07 ± ·20	1·143 ± ·054	4·70 ± ·22
12	7·237 ± ·036	·920 ± ·025	12·72 ± ·36	2·060 ± ·098	9·49 ± ·46
13	7·693 ± ·047	·856 ± ·033	11·13 ± ·46	1·842 ± ·124	7·98 ± ·54
14	6·440 ± ·041	·752 ± ·029	11·69 ± ·46	1·816 ± ·122	9·40 ± ·64
15	7·893 ± ·033	·842 ± ·023	10·67 ± ·30	1·737 ± ·083	7·34 ± ·35
16	7·226 ± ·040	·965 ± ·028	13·35 ± ·39	2·327 ± ·117	10·74 ± ·55
17	7·873 ± ·035	·908 ± ·025	11·53 ± ·32	2·144 ± ·102	9·08 ± ·44
18	7·467 ± ·039	·708 + ·028	9·49 ± ·37	1·414 ± ·095	6·31 ± ·43
19	7·181 ± ·038	·936 ± ·027	13·03 ± ·38	2·007 ± ·101	9·31 ± ·47
20	6·658 ± ·045	1·025 ± ·032	15·39 ± ·48	2·091 ± ·111	10·47 ± ·56
21	7·453 ± ·043	·788 ± ·031	10·58 ± ·42	1·775 ± ·120	7·94 ± ·54
22	7·570 ± ·038	·732 ± ·027	9·67 ± ·36	1·557 ± ·100	6·86 ± ·44
23	6·060 ± ·045	·818 ± ·032	13·50 ± ·53	1·894 ± ·128	10·42 ± ·71
24	7·128 ± ·045	·888 ± ·032	12·46 ± ·45	1·907 ± ·117	8·92 ± ·55
25	8·154 ± ·061	1·256 ± ·043	15·40 ± ·54	3·329 ± ·198	13·61 ± ·82
26	6·953 ± ·046	·843 ± ·033	12·13 ± ·18	1·844 ± ·124	8·84 ± ·60
27	7·067 ± ·043	·894 ± ·030	12·66 ± ·44	2·128 ± ·126	10·04 ± ·60
29	6·663 ± ·030	·781 ± ·021	11·72 ± ·33	1·763 ± ·084	8·82 ± ·42
30	5·917 ± ·029	·659 ± ·020	11·14 ± ·35	1·220 ± ·065	6·87 ± ·37
31	7·237 ± ·030	·783 ± ·022	10·82 ± ·30	1·687 ± ·080	7·77 ± ·37
32	7·713 ± ·035	·723 ± ·025	9·38 ± ·32	1·557 ± ·092	6·73 ± ·40
33	6·648 ± ·035	·856 ± ·025	12·87 ± ·38	1·946 ± ·098	9·76 ± ·49
34	7·053 ± ·046	·838 ± ·033	11·88 ± ·47	2·082 ± ·140	9·84 ± ·67
35	7·490 ± ·029	·741 ± ·020	9·90 ± ·27	1·513 ± ·072	6·73 ± ·32
37	8·337 ± ·050	1·292 ± ·036	15·50 ± ·44	3·254 ± ·155	13·01 ± ·63
38	6·020 ± ·025	·648 ± ·018	10·76 ± ·30	1·489 ± ·071	8·24 ± ·40
39	7·933 ± ·041	·668 ± ·029	8·42 ± ·37	1·470 ± ·111	6·17 ± ·47
40	7·118 ± ·043	·884 ± ·030	12·42 ± ·43	1·973 ± ·117	9·24 ± ·55

the former. The frequencies and constants for total ovules per ovary are conveniently tabled here for a later problem.

For purposes of comparison we may now take the differences between the three series. We use:

B—A, or the constant for the developing ovules minus the constant for the eliminated ovaries.

C—B, or the constant for the matured fruits minus the constant for those in early stages of development.

C—A, or the constant for the matured fruits minus the constant for the eliminated ovaries.

The three comparisons are made in Table IV.

TABLE II.

Physical Constants for Series B.

Number of Shrub	Ovules per Locule			Ovules per Ovary	
	Average and Probable Error	Standard Deviation and Probable Error	Coefficient of Variation	Standard Deviation and Probable Error	Coefficient of Variation
11	8·157 ± ·026	·672 ± ·018	8·24 ± ·23	1·300 ± ·062	5·31 ± ·25
12	6·913 ± ·038	·990 ± ·027	14·32 ± ·40	2·427 ± ·011	11·70 ± ·57
13	7·827 ± ·024	·624 ± ·017	7·97 ± ·22	1·109 ± ·053	4·72 ± ·23
14	6·573 ± ·050	·912 ± ·035	13·87 ± ·55	2·289 ± ·154	11·61 ± ·79
15	8·087 ± ·031	·786 ± ·022	9·73 ± ·27	1·487 ± ·071	6·13 ± ·29
16	7·141 ± ·043	1·038 ± ·030	14·53 ± ·43	2·608 ± ·131	12·17 ± ·62
17	8·137 ± ·033	·847 ± ·023	10·41 ± ·29	1·795 ± ·086	7·35 ± ·35
18	7·080 ± ·035	·906 ± ·025	12·79 ± ·36	2·168 ± ·103	10·21 ± ·49
19	7·246 ± ·043	·993 ± ·031	13·70 ± ·43	2·257 ± ·120	10·38 ± ·56
20	7·107 ± ·037	·915 ± ·027	12·87 ± ·38	2·049 ± ·103	9·61 ± ·49
21	7·772 ± ·025	·604 ± ·018	7·77 ± ·23	1·366 ± ·071	5·86 ± ·30
22	7·620 ± ·031	·640 ± ·022	8·40 ± ·29	1·252 ± ·074	5·48 ± ·32
23	6·140 ± ·040	·721 ± ·028	11·75 ± ·46	3·544 ± ·239	19·24 ± ·34
24	7·026 ± ·046	·947 ± ·032	13·48 ± ·47	2·269 ± ·134	10·77 ± ·64
25	8·883 ± ·046	1·190 ± ·033	13·40 ± ·38	2·517 ± ·120	9·45 ± ·45
26	7·093 ± ·049	·882 ± ·034	12·44 ± ·49	1·929 ± ·130	9·07 ± ·62
27	7·432 ± ·042	·960 ± ·030	12·91 ± ·40	2·135 ± ·114	9·58 ± ·52
29	6·383 ± ·034	·874 ± ·024	13·69 ± ·38	2·007 ± ·096	10·48 ± ·50
30	6·047 ± ·032	·760 ± ·023	12·57 ± ·38	1·674 ± ·087	9·23 ± ·48
31	6·860 ± ·032	·833 ± ·023	12·14 ± ·34	1·960 ± ·093	9·53 ± ·46
32	7·677 ± ·029	·738 ± ·020	9·62 ± ·27	1·729 ± ·082	7·51 ± ·36
33	6·357 ± ·036	·864 ± ·026	13·59 ± ·41	1·896 ± ·098	9·94 ± ·52
34	7·089 ± ·042	1·036 ± ·030	14·61 ± ·43	2·322 ± ·117	10·92 ± ·55
35	7·730 ± ·022	·557 ± ·015	7·21 ± ·20	1·046 ± ·050	4·51 ± ·22
37	8·763 ± ·041	1·046 ± ·029	11·94 ± ·33	2·442 ± ·116	9·29 ± ·45
38	5·833 ± ·026	·668 ± ·018	11·45 ± ·32	1·411 ± ·067	8·06 ± ·39
39	8·240 ± ·027	·699 ± ·019	8·49 ± ·24	1·429 ± ·068	5·78 ± ·28
40	7·780 ± ·034	·874 ± ·024	11·24 ± ·31	2·006 ± ·096	8·59 ± ·41

The means of the eliminated ovaries will be sensibly identical with those of the developing series if there is no selective elimination. If there is a tendency to the elimination of those with high numbers of ovules, all three of these differences will have the negative sign, while, if it is the ovaries with the lower number of ovules which are the more likely to be eliminated, the difference will carry the positive sign. These statements apply only to such differences as are greater than those liable to arise through the errors of random sampling. When differences are slight, the errors due to this source may be considerable enough to change the sign of a constant from positive to negative. The probable errors of the differences are also given, and to facilitate comparison the ratios of the differences to their probable error, are also tabulated.

Before considering the ratios of the individual differences to their probable errors we may examine the table as a whole, basing our conclusions concerning the presence or the absence of a selective elimination merely upon the signs of the differences in the mean number of ovules per locule.

If there be no selective elimination whatever, the means calculated from the three collections should be sensibly identical—that is to say, identical except for the differences due to the errors of random sampling. If our series of shrubs were very large we should expect half of the differences between the values of any two constants to be negative and half to be positive, if these differences were due merely to chance differences in the gathering of the fruits composing the samples, just as we would expect 50 per cent. heads and 50 per cent. tails in tossing a coin a few hundred times. We have only 28 throws of our coin, to use our figure, and so we cannot expect to get exactly 14 of each sign, but we ought not to get very wide divergences from these numbers unless there is some biological factor at work to modify the proportions.

That there must be some such factor is quite evident from a casual examination of the tables. Taking first the differences between the eliminated ovaries and those which are continuing their development, we find that the mean number of

TABLE III.

Physical Constants for Series C.

Number of Shrub	Ovules per Locule			Ovules per Ovary	
	Average and Probable Error	Standard Deviation and Probable Error	Coefficient of Variation	Standard Deviation and Probable Error	Coefficient of Variation
11	8·360 ± ·030	·760 ± ·021	9·09 ± ·25	1·804 ± ·086	7·19 ± ·34
12	7·480 ± ·031	·793 ± ·022	10·61 ± ·28	1·657 ± ·079	7·38 ± ·35
13	6·757 ± ·036	·915 ± ·025	13·54 ± ·38	1·891 ± ·090	9·33 ± ·45
14	6·588 ± ·035	·746 ± ·025	11·32 ± ·38	1·664 ± ·096	8·42 ± ·49
15	8·713 ± ·032	·836 ± ·023	9·59 ± ·27	1·980 ± ·094	7·57 ± ·36
16	8·190 ± ·033	·860 ± ·024	10·51 ± ·29	2·127 ± ·101	8·66 ± ·42
17	8·730 ± ·041	1·051 ± ·029	12·03 ± ·34	2·671 ± ·127	10·20 ± ·49
18	7·587 ± ·029	·754 ± ·021	9·94 ± ·28	1·861 ± ·089	8·18 ± ·39
19	7·850 ± ·027	·689 ± ·019	8·77 ± ·24	1·639 ± ·078	6·96 ± ·33
20	7·643 ± ·028	·709 ± ·019	9·28 ± ·26	1·538 ± ·073	6·71 ± ·32
21	8·067 ± ·025	·654 ± ·018	8·11 ± ·22	1·544 ± ·074	6·38 ± ·31
22	7·903 ± ·020	·511 ± ·014	6·47 ± ·18	1·151 ± ·055	4·86 ± ·23
23	6·261 ± ·039	·777 ± ·028	12·41 ± ·45	1·644 ± ·101	8·75 ± ·54
24	7·537 ± ·028	·713 ± ·020	9·46 ± ·26	1·378 ± ·066	6·09 ± ·29
25	9·597 ± ·077	1·990 ± ·055	20·74 ± ·59	1·930 ± ·092	6·70 ± ·32
26	7·618 ± ·030	·668 ± ·021	8·77 ± ·28	1·429 ± ·078	6·25 ± ·34
27	8·103 ± ·027	·688 ± ·019	8·49 ± ·24	1·347 ± ·064	5·54 ± ·26
29	6·917 ± ·031	·802 ± ·022	11·59 ± ·32	1·763 ± ·084	8·49 ± ·41
30	6·387 ± ·026	·666 ± ·018	10·43 ± ·29	1·294 ± ·062	6·75 ± ·32
31	7·463 ± ·029	·750 ± ·021	10·05 ± ·28	1·679 ± ·080	7·50 ± ·36
32	7·810 ± ·026	·664 ± ·018	8·50 ± ·24	1·471 ± ·070	6·28 ± ·30
33	6·833 ± ·030	·783 ± ·022	11·45 ± ·32	1·775 ± ·085	8·66 ± ·42
34	7·647 ± ·029	·740 ± ·020	9·68 ± ·27	1·541 ± ·073	6·72 ± ·32
35	7·830 ± ·022	·578 ± ·016	7·38 ± ·20	1·127 ± ·054	4·80 ± ·23
37	8·533 ± ·040	1·021 ± ·028	11·97 ± ·33	2·112 ± ·101	8·25 ± ·40
38	7·003 ± ·035	·908 ± ·025	12·96 ± ·36	2·343 ± ·112	11·15 ± ·53
39	3·290 ± ·029	·743 ± ·020	8·97 ± ·25	1·540 ± ·073	6·19 ± ·30
40	8·233 ± ·028	·721 ± ·020	8·75 ± ·24	1·546 ± ·074	6·26 ± ·30

TABLE IV.

Difference between Mean Number of Ovules per Locule for three Series of Ovaries.

Number of Shrub	B—A		C—B		C—A	
	Difference and Probable Error	Diff. / E diff.	Difference and Probable Error	Diff. / E diff.	Difference and Probable Error	Diff. / E diff.
11	·040 ± ·034	1·16	·203 ± ·039	5·14	·243 ± ·037	6·55
12	− ·323 ± ·053	6·15	·567 ± ·049	11·47	·243 ± ·047	5·14
13	·133 ± ·053	2·51	− 1·070 ± ·043	24·82	− ·937 ± ·059	15·85
14	·133 ± ·065	2·05	·015 ± ·061	·24	·148 ± ·054	2·73
15	·193 ± ·045	4·31	·627 ± ·046	13·71	·820 ± ·046	17·75
16	− ·085 ± ·058	1·46	1·049 ± ·045	23·21	·964 ± ·052	18·58
17	·263 ± ·048	5·44	·593 ± ·052	11·30	·857 ± ·054	15·83
18	·333 ± ·053	6·34	− ·213 ± ·046	4·65	·120 ± ·049	2·46
19	·064 ± ·058	1·11	·604 ± ·051	11·89	·669 ± ·047	14·29
20	·449 ± ·058	7·70	·536 ± ·047	11·50	·985 ± ·052	18·83
21	·319 ± ·050	6·35	·294 ± ·035	8·50	·613 ± ·050	12·19
22	·051 ± ·049	1·03	·283 ± ·037	7·68	·334 ± ·043	7·70
23	·080 ± ·060	1·33	·121 ± ·056	2·17	·201 ± ·060	3·37
24	− ·102 ± ·064	1·60	·511 ± ·054	9·53	·409 ± ·053	7·77
25	·729 ± ·076	9·55	·713 ± ·090	7·90	1·443 ± ·098	14·66
26	·140 ± ·074	1·88	·525 ± ·057	9·21	·665 ± ·055	12·03
27	·367 ± ·060	6·10	·670 ± ·050	13·51	1·037 ± ·051	20·40
29	− ·280 ± ·033	8·51	·533 ± ·046	11·57	·253 ± ·044	5·81
30	·130 ± ·043	3·03	·340 ± ·041	8·24	·470 ± ·039	12·18
31	− ·377 ± ·044	8·46	·603 ± ·044	13·84	·227 ± ·042	5·37
32	− ·036 ± ·045	·80	·133 ± ·039	3·45	·097 ± ·043	2·24
33	− ·291 ± ·039	7·37	·476 ± ·048	10·01	·185 ± ·046	3·98
34	·036 ± ·063	·57	·558 ± ·051	10·87	·593 ± ·054	10·93
35	·240 ± ·036	6·65	·100 ± ·031	3·19	·340 ± ·037	9·29
37	·427 ± ·065	6·59	− ·230 ± ·057	4·04	·197 ± ·065	3·03
38	− ·187 ± ·036	5·16	1·170 ± ·044	26·65	·983 ± ·043	22·66
39	·307 ± ·049	6·22	·050 ± ·040	1·26	·357 ± ·050	7·10
40	·662 ± ·055	12·13	·453 ± ·044	10·26	1·115 ± ·051	21·83

ovules is higher in the developing series in 20 cases and higher in the eliminated series in 8 cases. This is a deviation of only 6 shrubs from the theoretical 14 of each class and hence cannot be regarded as very strongly significant. So far as the evidence goes, it indicates that ovaries bearing a smaller number of ovules are less likely to continue their development towards maturity than those having a larger number.

Even if there were a strong selective elimination of ovaries with a smaller number of ovules per locule, we would not expect to detect the full intensity of such a selective elimination in a comparison of the two series A and B, for the simple reason that we are comparing an eliminated series of ovaries with a series in which further elimination is to be expected.

If it is true that the elimination is not yet complete in the B series, and that the failure of ovaries to develop is not random, but selective, we would expect

to find differences between the B series and the matured fruits—series C—of the same prevailing sign as between the series A and B.

The results obtained by taking the difference C—B bear out in a most striking manner the considerations drawn from the differences B—A, for only 3 out of the 28 differences bear the negative sign!

Finally the most critical comparison is that between the sample of eliminated ovaries and the fruits which reach maturity, i.e. C—A. The results bear out fully the conclusions drawn from the two foregoing comparisons, for only *one* of the 28 differences has the negative sign.

If any biologist wishes to test the number of cases in which twenty-seven positive and one negative sign would be expected according to purely chance causes, he might take 28 coins and toss them, noting the number of heads and tails he gets in each toss in order to determine the number of times that he will find 27 heads and 1 tail in a few hundred tosses. Mathematical theory says it will be about 46 times in 100,000,000.

The reader may take these chances for what he considers them worth. Personally I feel that the evidence is rather strongly in favour of the conclusion that there is a selective elimination of ovaries and that on the average it is those with the larger number of ovules which reach maturity.

In Diagram 1 I have tried to make patent to the eye the difference between the eliminated and the matured ovaries for each of the 28 individuals and for all of the material examined. Here the circles show the mean of the eliminated ovaries and the solid dots the mean of those which reach maturity. The length of

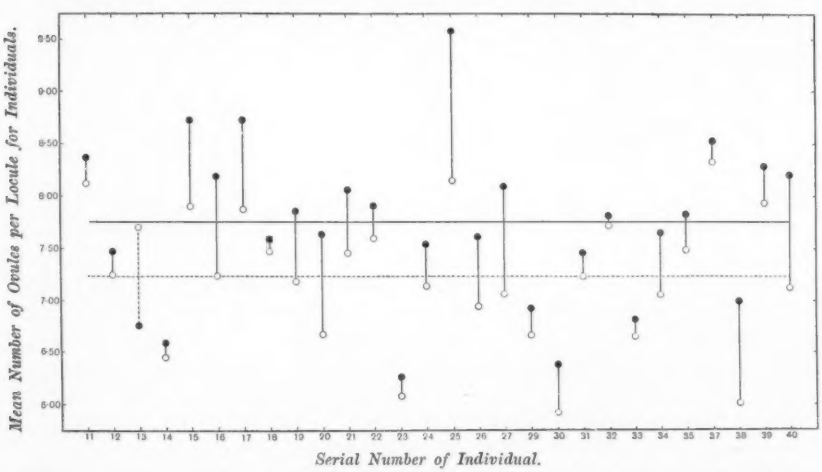


Diagram 1. Showing difference in mean of eliminated and matured ovaries for the 28 individuals and for the population. Circles = mean of eliminated locules for individuals; solid dots = mean of matured locules for individuals. Broken transverse line = mean of all eliminated locules; solid transverse line = mean of all matured locules.

the line connecting these shows the difference between the two samples for each individual. The transverse lines show the position of the means for the whole population, the broken one being for the eliminated and the continuous one for the matured ovaries.

The diagram shows one further point—the great difference in the means of the individuals. The two transverse lines may be regarded as in a way smoothing these irregularities. The solid line is really the mean position of the solid dots on the scale while the broken line shows the mean position of the circles, each individual being weighted with the number of ovaries which it furnished. I think the differences shown on the diagram are conspicuous enough to carry considerable weight.

Let us now consider the significance of the constants from the individual trees. We shall not expect this to be very high as judged by a comparison of the probable errors, because of the smallness of the samples of material taken. Unfortunately the excessive labour of the dissection and the counting of the ovaries, even in the mature state, precluded the examination of more extensive collections.

The trustworthiness of the difference between any two constants is indicated by its probable error. The probable error of the difference between any two uncorrelated* constants, say x and y, may be calculated by the formula

$$E(x-y) = \sqrt{Ex^2 + Ey^2}.$$

Conventionally, a constant is not considered significant unless it is 2·5 or more times its probable error. This is the degree of divergence from 0 demanded in the following discussion. For convenience of comparison the ratio

$$\frac{\text{Difference}}{\text{Probable Error of Difference}}$$

has been tabled for the three comparisons for each tree (see the third column, Table IV.).

For the comparison B—A it appears that there are 13 significantly positive and 5 significantly negative differences. For the comparison C—B there are 22 significantly positive and 3 significantly negative differences.

Finally for the most critical test of selective elimination, C—A, there are 25 significantly positive and 1 significantly negative differences.

The thing which strikes one particularly about these conclusions is the high number of cases in which the differences between the constants are significant with

* I am afraid that here our constants are not perfectly independent. As Professor Pearson suggested to me, this is certainly not true where the size of samples is at all large as compared with the total number of ovaries produced, for a random excess in a first sample would be associated with a random deficit in the second.

I do not believe that all of my samples combined equal one-tenth of the total ovaries produced, and I suspect that the influence of the correlation of the samples would not be large, but I have no statistical proof of this. Indeed in a form like *Staphylea* it would be difficult to get suitable data for the correction of the probable error of the differences in two constants, even if the formulae were available. I think these constants affect in no way the validity of the conclusions drawn in this paper, but they may render some of the probable errors somewhat questionable.

respect to their probable error. In 18, 25 and 26 cases out of the 28 the differences exceed 2·5 times their probable errors, and often reach many times their probable error. I think relationships of significantly positive to significantly negative differences, such as 13 : 5, 22 : 3, 25 : 1 can leave no doubt in the mind of the most sceptical that there is a very potent selective elimination, by which the mean number of ovules per ovary is very materially increased.

The combination of the material from all the individuals gives the results for the differences in the mean number of ovules per locule presented in Table V.

TABLE V.

Comparison of Means for total Material.

Comparison	Difference and Probable Error	Ratio of Difference to its Probable Error	Percentage Increase in Mean
B—A	·1426 ± ·0129	11·054	1·971
C –B	·3693 ± ·0120	30·775	5·005
C—A	·5119 ± ·0121	42·306	7·075

Here the absolute differences are given in the first column, and the percentage increase of the higher over the lower mean in the last column. The results emphasize the conclusions drawn from the constants for the individual trees.

The difference between the eliminated sample and the sample of ovaries remaining on the tree when they were taken is only about 2 per cent. of the former, but the difference is about 11 times its probable error and thus certainly significant.

Doubtless the very slight divergence between the series A and B is due to the presence of a large number of ovaries in B that would a few days later fall from the tree and be classed with the A's, for we note a much greater increase from B to C than from A to B, amounting to no less than 5 per cent.

As is to be expected the difference between the eliminated and the matured fruits is greatest of all, the latter having about 7 per cent. more ovules per locule than the former.

The reader will note that in the last two comparisons the differences are 31 and 42 times their probable errors.

As a further demonstration of the differences in number of ovules per locule or per fruit in the three series I give (Diagram 2), the data in the form of an integral polygon in which the frequency on each grade is the sum of its own frequency and that of all lower grades. This form is one of the most convenient for purposes of comparison.

It is unnecessary to table the actual frequencies for they can easily be computed by anyone who requires them from the totals of our tables. The percentage

frequencies for number of ovules per locule are given in Table VI, and for total number of ovules per fruit in Table VII. This method of presenting the data shows very clearly the selective elimination. For instance, in the comparison of

TABLE VI.

Percentile Frequencies of Ovules per Locule in three Series.

Number of Ovules	A	B	C
2	—	·01	—
3	·03	·03	—
4	·22	·35	·04
5	4·25	4·27	·74
6	27·86	25·19	15·78
7	54·64	48·30	31·39
8	91·50	88·00	59·79
9	98·01	96·59	93·11
10	99·94	99·50	99·78
11	100·00	99·93	99·99
12	—	100·00	100·00

TABLE VII.

Percentile Frequencies of total Ovules per Ovary in three Series.

Number of Ovules	A	B	C
11	—	·04	—
12	·05	·04	·04
13	·05	·12	·04
14	·19	·24	·04
15	·76	·77	·04
16	2·10	2·60	·18
17	6·01	6·33	·96
18	16·13	15·58	6·92
19	23·72	22·60	11·69
20	33·17	30·30	18·19
21	43·44	38·54	24·15
22	57·42	47·99	33·28
23	71·88	62·27	45·12
24	88·50	83·04	76·18
25	93·75	89·86	84·95
26	96·18	93·87	89·79
27	97·57	95·94	92·71
28	98·90	97·73	94·75
29	99·67	99·03	97·15
30	99·90	99·80	99·63
31	100·00	99·96	99·96
32	—	99·96	100·00
33	—	100·00	—

number of ovules per locule we note that locules with 7 ovules or less form 54·6 per cent. of the collection, while in the collection of matured fruits they form only 31·4 per cent. In all three series we are dealing with over 2000 fruits, or over 6000 individual locules, so the results can be looked upon with considerable confidence.

Diagrams 2 and 3 show these results graphically.

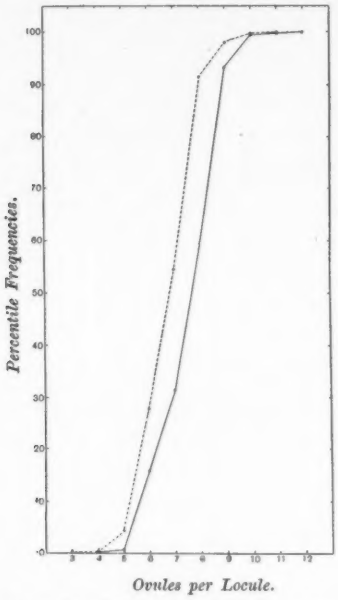


DIAGRAM 2. Percentile curve showing frequency of different numbers of ovules per locule in the eliminated series and the series developing to maturity. Dots and broken lines = Eliminated series; dots and firm lines = Matured series.

Problem 2. Is there a difference in the variability in number of ovules per locule and per fruit in the eliminated ovaries and those which develop to maturity?

The results of the preceding section indicate clearly that the chances of an ovary developing to maturity are not independent of its number of ovules but that those with the smaller numbers are apt to be eliminated.

It is quite thinkable that a strong selective elimination might exist without bringing about any sensible difference in the mean values of a character in the population before and after elimination. If the elimination were not of the larger or the smaller fruits as measured by the number of ovules per locule or per fruit, but of *both* largest and smallest fruits, i.e. of the *extreme* variates of both kinds, the

mean values of the number of ovules in the matured fruits might not differ sensibly from that for the original ovaries, but the variability in number would be reduced.

On the other hand it is not impossible that a selective elimination might effect chiefly the modal region of the curve of distribution, materially flattening it off and so increasing the variability as measured by the standard deviation. To the biologist this seems theoretically less probable than the elimination of either the largest or the smallest ovaries, or the extreme variates of both types, but of course all possibilities must be tested out on the basis of actual data.

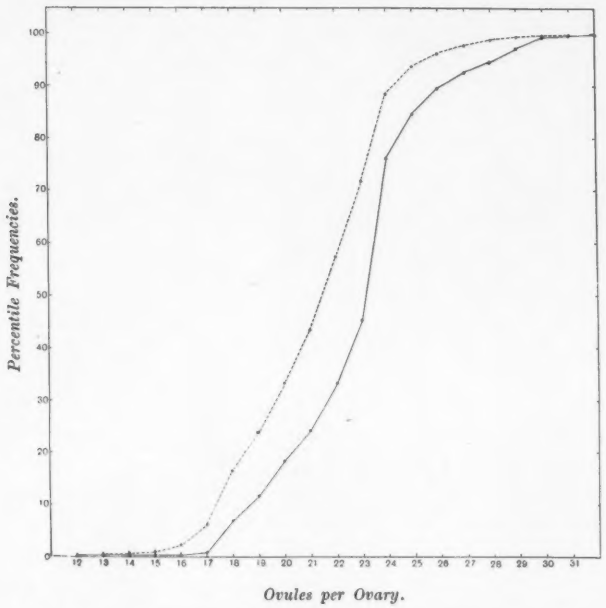


DIAGRAM 3. Percentile curve showing frequency of different numbers of ovules per ovary in the eliminated series and the series developing to maturity. Dots and broken line = Eliminated series; dots and firm line = Matured series.

In comparing the means of the eliminated and uneliminated series, it was only necessary to determine the values for either ovules per locule or ovules per fruit, since one is necessarily three times the other. For the variabilities, however, it seems best to examine the standard deviation for both ovules per locule and total ovules per ovary, since one cannot readily be got from the other.

We are dealing with integral variates, and hence Sheppard's modification for the second moment was not applied.

Standard Deviations.

First, let us deal with absolute variabilities. The standard deviations for the number of ovules per locule in the three collections are given in Tables I—III, and are compared in Table VIII.

TABLE VIII.

Comparison of Standard Deviations of Number of Ovules per Locule in the three Series.

Number of Shrub	B—A		C—B		C—A	
	Difference and Probable Error	Diff. / E diff.	Difference and Probable Error	Diff. / E diff.	Difference and Probable Error	Diff. / E diff.
11	·098 ± ·024	4·04	·088 ± ·028	3·14	·186 ± ·026	7·09
12	·069 ± ·037	1·87	− ·196 ± ·035	5·61	− ·127 ± ·033	3·79
13	− ·232 ± ·037	6·19	·291 ± ·030	9·54	·059 ± ·042	1·40
14	·159 ± ·046	3·46	− ·166 ± ·043	3·83	− ·007 ± ·038	·17
15	− ·156 ± ·032	1·75	·049 ± ·032	1·56	− ·006 ± ·033	·20
16	·073 ± ·041	1·78	− ·177 ± ·038	4·63	− ·104 ± ·037	2·84
17	− ·061 ± ·034	1·78	·204 ± ·037	5·50	·143 ± ·038	3·73
18	·197 ± ·037	5·31	− ·152 ± ·032	4·68	·046 ± ·035	1·32
19	·057 ± ·041	1·50	− ·304 ± ·036	8·45	− ·247 ± ·033	7·44
20	− ·110 ± ·041	2·67	− ·205 ± ·033	6·22	− ·315 ± ·037	8·50
21	− ·185 ± ·036	5·18	·051 ± ·025	2·02	− ·134 ± ·036	3·76
22	− ·092 ± ·035	2·62	− ·129 ± ·026	4·97	− ·221 ± ·031	7·22
23	− ·097 ± ·042	2·28	·056 ± ·039	1·42	− ·041 ± ·042	·97
24	·059 ± ·045	1·31	− ·234 ± ·038	6·20	− ·175 ± ·037	4·71
25	− ·065 ± ·054	1·21	·800 ± ·064	12·51	·734 ± ·070	10·55
26	·039 ± ·047	·81	− ·214 ± ·040	5·31	− ·175 ± ·039	4·49
27	·065 ± ·042	1·53	− ·272 ± ·035	7·74	− ·207 ± ·036	5·75
29	·092 ± ·032	2·86	− ·072 ± ·033	2·20	·020 ± ·031	·66
30	·101 ± ·030	3·32	− ·094 ± ·029	3·24	·007 ± ·028	·25
31	·050 ± ·031	1·57	− ·083 ± ·031	2·69	− ·033 ± ·030	1·11
32	·015 ± ·032	·47	− ·074 ± ·027	2·72	− ·059 ± ·031	1·93
33	·008 ± ·036	·23	− ·081 ± ·034	2·42	− ·073 ± ·033	2·23
34	·198 ± ·044	4·67	− ·296 ± ·036	8·13	− ·097 ± ·038	2·54
35	− ·184 ± ·025	7·23	·021 ± ·022	·96	− ·163 ± ·026	6·33
37	− ·246 ± ·046	5·37	− ·025 ± ·040	·62	− ·271 ± ·045	5·98
38	·020 ± ·026	·78	·240 ± ·031	7·74	·260 ± ·031	8·47
39	·031 ± ·035	·90	·044 ± ·028	1·56	·075 ± ·036	2·12
40	− ·010 ± ·039	·25	− ·154 ± ·031	4·95	− ·163 ± ·036	4·53

The differences are taken in the same way as for the average number of ovules per locule, i.e. B—A, C—B, and C—A. If selection tends to reduce variability the negative sign should predominate in the differences; if selection tends to increase variability the differences with the positive sign should be the more frequent.

To render the comparisons more critical, the probable errors of the standard deviations and the probable errors of their differences have been tabled. The ratio of the difference to its probable error is also given in the same manner as for the means.

Studying the tables first for the prevalence of positive and negative signs we find the results:

B—A, 11 negative and 17 positive differences,

C—B, 18 negative and 10 positive differences,

C—A, 19 negative and 9 positive differences.

Taking up next the significance of the differences for the constants for individuals as measured by their ratio to their probable error, we find results as follows.

For the comparison B—A only 12 of the 28 differences are significant as compared with 18 out of 28 in the case of the means. Of these 6 are positive and 6 negative in sign. The difference C—B shows 20 significant constants, of which 5 are positive and 15 negative. C—A shows 17 significant constants, of which 4 are positive and 13 negative. The first comparison indicates that selective elimination has no influence upon variability. The second and third both suggest that the variability is sensibly less after elimination has taken place.

The differences in the variabilities for the lumped collections may now be examined in Table IX.

TABLE IX.

Comparison	Difference and Probable Error	Diff. / *E* diff.
B—A	+·0756 ± ·0091	8·31
C—B	−·0938 ± ·0086	10·91
C—A	−·0182 ± ·0081	2·25

At first these results puzzled me very greatly; and the following interpretation is offered only tentatively and as a working hypothesis for further investigation.

The collection B probably comes the nearest to representing the original population of ovaries of any that we have, since it still contains a high percentage of ovaries which will be eliminated later*.

Now it is quite clear that if the ovaries with the *lower* numbers of ovules per locule are eliminated, their variability will be expected to be less than that of the original population. Likewise if the ovaries with the *higher* numbers of ovaries develop to maturity it is clear that *their* variability also should be less than that of the population as a whole. In short, both the samples of eliminated ovaries and the samples of matured fruits represent the more extreme portions of the curve of variation. If this be true, we should expect to find the variability of B greater than that of A or C, while an inevitable consequence of this condition of affairs

* I am sorry not to be able to give definite quantitative data bearing upon this point, but this is quite out of the question in a ligneous plant like *Staphylea*, where the number of flowers produced is immensely large and where it is impossible to determine with certainty from the mature inflorescence the number of flowers originally produced.

would be that the difference in the variability of C and A would be less than that of either of the other two differences.

This is precisely the condition which we find; for the total material the differences B—A and C—B are in opposite senses and clearly significant with regard to the probable errors of their differences, while the difference between C and A cannot safely be regarded as significant.

Returning now to the results for the individual shrubs, we note that so far as the signs give evidence it is that A and C are less variable than B, just as our hypothesis demands. For the differences for individuals which are significant with regard to their probable errors, there are 6 (A < B) and 6 (A > B). For C—B there are 5 (C > B) and 15 (C < B). Finally C—A gives 4 (C > A) and 13 (C < A). These results contain only a partial substantiation for our hypothesis, but no evidence directly against it.

Only further research will show whether these conditions are generally found.

Take next the results for total ovules per ovary. The comparison is made in Table X.

TABLE X. *Comparison of Standard Deviations of total Ovules per Ovary in the three Series.*

Number of Shrub	B—A		C—B		C—A	
	Difference and Probable Error	Diff. / *E* diff.	Difference and Probable Error	Diff. / *E* diff.	Difference and Probable Error	Diff. / *E* diff.
11	·156 ± ·082	1·89	·504 ± ·106	4·76	·660 ± ·101	6·51
12	·367 ± ·099	3·71	− ·770 ± ·080	9·65	− ·040 ± ·126	·32
13	− ·733 ± ·135	5·43	·782 ± ·104	7·49	·005 ± ·153	·03
14	·473 ± ·197	2·40	− ·626 ± ·182	3·43	− ·015 ± ·156	·10
15	− ·250 ± ·109	2·30	·493 ± ·118	4·18	·243 ± ·125	1·94
16	·280 ± ·175	1·60	− ·480 ± ·165	2·90	− ·020 ± ·155	·13
17	− ·035 ± ·133	·26	·876 ± ·153	5·72	·053 ± ·163	·32
18	·754 ± ·141	5·36	− ·308 ± ·136	2·26	·045 ± ·130	·34
19	·250 ± ·157	1·59	− ·617 ± ·143	4·31	− ·037 ± ·127	·29
20	− ·043 ± ·152	·28	− ·511 ± ·126	4·05	− ·055 ± ·133	·41
21	− ·409 ± ·139	2·94	·178 ± ·102	1·74	− ·231 ± ·140	1·65
22	− ·305 ± ·124	2·45	− ·100 ± ·192	1·09	− ·041 ± ·114	·35
23	1·650 ± ·271	6·09	− 1·900 ± ·260	7·32	− ·250 ± ·163	1·53
24	·362 ± ·178	2·03	− ·891 ± ·149	5·97	− ·530 ± ·134	3·94
25	− ·811 ± ·231	3·52	− ·587 ± ·151	3·88	− 1·399 ± ·217	6·44
27	·085 ± ·180	·47	− ·500 ± ·152	3·30	− ·041 ± ·147	·28
28	·007 ± ·170	·04	− ·789 ± ·130	6·05	− ·078 ± ·141	·55
29	·243 ± ·127	1·91	− ·244 ± ·127	1·92	− ·000 ± ·119	·00
30	·454 ± ·108	4·20	− ·380 ± ·106	3·58	·074 ± ·090	·83
31	·273 ± ·123	2·22	− ·282 ± ·123	2·29	·008 ± ·114	·07
32	·172 ± ·124	1·39	− ·257 ± ·108	2·38	− ·086 ± ·116	·74
33	− ·050 ± ·138	·36	− ·121 ± ·129	·94	− ·171 ± ·129	1·33
34	·241 ± ·183	1·32	− ·781 ± ·138	5·67	− ·540 ± ·158	3·41
35	− ·467 ± ·188	5·32	·081 ± ·173	1·10	− ·386 ± ·090	4·29
37	− ·812 ± ·194	4·19	− ·331 ± ·154	2·15	− 1·142 ± ·185	6·18
38	− ·078 ± ·098	·80	·932 ± ·130	7·15	·854 ± ·132	6·46
39	− ·041 ± ·130	·31	·112 ± ·100	1·12	·071 ± ·133	·53
40	·033 ± ·151	·22	− ·460 ± ·121	3·81	− ·427 ± ·138	3·82

The frequencies of positive and negative differences, both for all differences and significant differences, for the individuals are shown in Table XI.

TABLE XI.

Comparison	All Differences		Significant Differences	
	Positive	Negative	Positive	Negative
B—A	16	12	5	5
C—B	8	20	4	13
C—A	9	19	2	6

The comparison of the variabilities calculated from the ovaries for all 28 trees is made in Table XII.

TABLE XII.

Comparison	Difference and Probable Error	$\frac{\text{Diff.}}{E\ \text{diff.}}$	Percentage Increase or Decrease*
B—A	$+ \cdot 2630 \pm \cdot 0410$	6·41	+9·49
C—B	$- \cdot 2235 \pm \cdot 0389$	5·75	−7·37
C—A	$+ \cdot 0395 \pm \cdot 0387$	1·02	+1·43

From these tables the same conclusion is to be drawn as from that for the comparison for the variability of the number of ovules per locule, namely that both the eliminated ovaries, A, and the matured fruits, C, show a less variability than the general population from which a portion of the ovaries have been eliminated. These results are very clear in the comparison for the totals of the three series; they are not so evident on the small individual samples with their higher probable errors.

Coefficients of Variation.

In the preceding section it was shown that a selective elimination brings about a different mean number of ovules per locule or per fruit in the eliminated and the matured ovaries. The ovaries which develop to maturity have a higher average number of ovules than those which are eliminated. Possibly, then, our conclusions concerning variability may be open to question because of the fact that we are comparing the variabilities of samples with generally higher means with those with generally lower means. The absolute variability is generally roughly proportional to the mean value of a character.

* Increase or decrease from the eliminated to the uneliminated or from the less mature to the more mature.

To make our comparisons more comprehensive and to free them in so far as possible from the influence of the actual magnitudes, we may now make use of the relative variabilities as expressed by the coefficients of variation.

The differences between the three series for the coefficients of variation for both ovules per locule and ovules per fruit appear in Tables XIII and XIV*.

TABLE XIII.

Comparison of Coefficients of Variation of Number of Ovules per Locule in the three Series.

Number of Shrub	B—A		C—B		C—A	
	Coefficient of Variation	Diff. / E diff.	Coefficient of Variation	Diff. / E diff.	Coefficient of Variation	Diff. / E diff.
11	1·17 ± ·30	3·90	·85 ± ·34	2·50	2·02 ± ·32	6·31
12	1·60 ± ·54	2·96	−3·71 ± ·49	7·57	−2·11 ± ·46	4·58
13	−3·16 ± ·51	6·20	5·57 ± ·44	12·66	2·41 ± ·50	4·82
14	2·19 ± ·72	3·04	−2·55 ± ·67	3·81	− ·36 ± ·60	·72
15	− ·94 ± ·40	2·35	− ·13 ± ·38	·34	−1·08 ± ·40	2·70
16	1·82 ± ·58	3·14	−4·03 ± ·52	7·75	−2·84 ± ·49	5·80
17	−1·12 ± ·43	2·60	1·62 ± ·45	3·60	·50 ± ·47	1·06
18	3·31 ± ·52	6·36	−2·85 ± ·46	6·19	·45 ± ·46	·98
19	·67 ± ·57	1·17	−4·93 ± ·49	10·06	−4·26 ± ·45	9·47
20	2·52 ± ·61	4·13	−3·59 ± ·46	7·80	−6·11 ± ·55	11·11
21	−2·81 ± ·47	5·98	·34 ± ·32	1·06	−2·46 ± ·46	5·35
22	−1·27 ± ·46	2·76	−1·93 ± ·34	5·68	−3·20 ± ·40	8·00
23	−1·76 ± ·70	2·51	·67 ± ·64	1·05	−1·09 ± ·69	1·58
24	1·02 ± ·65	1·57	−4·02 ± ·54	7·44	−3·00 ± ·52	5·78
25	−2·00 ± ·66	3·03	7·34 ± ·70	10·48	5·33 ± ·80	6·66
26	·31 ± ·69	·45	3·66 ± ·56	6·53	−3·36 ± ·56	6·00
27	·25 ± ·59	·42	−4·42 ± ·47	9·40	−4·17 ± ·50	8·32
29	1·96 ± ·50	3·92	−2·10 ± ·50	4·20	− ·13 ± ·46	·28
30	1·43 ± ·52	2·67	−2·14 ± ·48	4·50	− ·71 ± ·45	1·58
31	1·32 ± ·45	2·93	−2·09 ± ·44	4·52	− ·77 ± ·41	1·88
32	·24 ± ·42	·57	−1·12 ± ·36	3·11	− ·88 ± ·40	2·20
33	·72 ± ·56	1·28	−2·14 ± ·52	4·11	−1·42 ± ·50	2·84
34	2·74 ± ·64	4·22	−4·93 ± ·51	9·67	−2·20 ± ·54	4·07
35	−2·69 ± ·32	8·41	·18 ± ·28	·64	−2·52 ± ·32	7·87
37	−3·56 ± ·55	6·47	·03 ± ·47	·06	−3·53 ± ·55	6·42
38	·69 ± ·44	1·57	1·51 ± ·48	3·14	2·20 ± ·47	4·68
39	·07 ± ·44	·16	·48 ± ·35	1·37	·55 ± ·45	1·22
40	−1·18 ± ·53	2·23	−2·49 ± ·39	6·38	−3·67 ± ·49	7·49

The results may be tersely summarized for the individuals in Table XV. Here the frequencies for positive and negative differences are shown, both for all the differences and for probably significant differences, i.e. for those which are 2·5 or more times their probable error.

* Naturally the significance of the sign as indicating the tendency of selection to increase or decrease variability is the same for V as for σ.

For the total material we have, for locules,

$$B—A = + \cdot 735 \pm \cdot 128,$$
$$C—B = - \cdot 961 \pm \cdot 116,$$
$$C—A = - \cdot 225 \pm \cdot 112,$$

and for total ovules per ovary

$$B—A = + \cdot 941 \pm \cdot 190,$$
$$C—B = - 1 \cdot 615 \pm \cdot 175,$$
$$C—A = - \cdot 674 \pm \cdot 176.$$

These results are consistent *throughout*, for both ovules per locule and total ovules per fruit. They show that, both for the series of individuals and the three grand totals, the relative as well as the absolute variability is higher in Series B than in Series A or C. The variability of the eliminated series, A, seems to be

TABLE XIV.

Comparison of Coefficients of Variation of Number of Ovules per Ovary in the three Series.

Number of Shrub	B—A		C—B		C—A	
	Coefficient of Variation	Diff. / *E* diff.	Coefficient of Variation	Diff. / *E* diff.	Coefficient of Variation	Diff. / *E* diff.
11	·61 ± ·33	1·85	1·88 ± ·42	4·47	2·50 ± ·40	6·25
12	2·21 ± ·73	3·03	− 4·32 ± ·67	6·44	− 2·11 ± ·58	3·63
13	− 3·26 ± ·59	5·52	4·61 ± ·50	9·22	1·35 ± ·70	1·93
14	2·21 ± ·02	2·17	− 3·19 ± ·93	3·43	− ·98 ± ·81	1·21
15	− 1·20 ± ·45	2·67	1·44 ± ·46	3·13	·24 ± ·50	·48
16	1·44 ± ·83	1·73	− 3·51 ± ·75	4·68	− 2·08 ± ·69	3·02
17	− 1·72 ± ·56	3·07	2·84 ± ·60	4·73	1·12 ± ·66	1·70
18	3·90 ± ·65	6·00	− 2·03 ± ·63	3·22	1·86 ± ·58	3·21
19	1·07 ± ·73	1·46	− 3·42 ± ·65	5·26	− 2·35 ± ·58	4·05
20	− ·86 ± ·74	1·62	− 2·90 ± ·58	5·00	− 3·76 ± ·64	5·87
21	− 2·08 ± ·62	3·35	·52 ± ·43	1·21	− 1·56 ± ·62	2·52
22	− 1·38 ± ·54	2·55	− ·62 ± ·39	1·59	− 2·00 ± ·50	3·33
23	8·82 ± 1·52	16·96	− 10·49 ± ·44	7·28	− 1·66 ± ·89	1·86
24	1·85 ± ·84	2·21	− 4·67 ± ·70	6·67	− 2·83 ± ·62	4·56
25	4·16 ± ·94	4·42	− 2·74 ± ·55	4·98	− 6·90 ± ·88	7·84
26	·23 ± ·86	·27	− 2·81 ± ·71	3·96	− 2·59 ± ·69	3·80
27	− ·46 ± ·79	·58	− 4·04 ± ·58	6·96	− 4·50 ± ·65	6·92
29	1·66 ± ·65	2·55	− 1·98 ± ·65	3·04	− ·33 ± ·58	·57
30	2·36 ± ·61	3·87	− 2·47 ± ·58	4·26	− ·12 ± ·49	·24
31	1·75 ± ·59	2·96	− 2·03 ± ·58	3·48	− ·27 ± ·52	·52
32	·78 ± ·54	1·44	− 1·23 ± ·47	2·62	− ·45 ± ·50	·90
33	·18 ± ·71	2·53	− 1·28 ± ·67	1·91	− 1·10 ± ·64	1·72
34	1·08 ± ·87	1·25	− 4·20 ± ·64	6·56	− 3·12 ± ·74	4·22
35	− 2·22 ± ·39	5·69	·29 ± ·32	·91	− 1·94 ± ·39	4·97
37	− 3·72 ± ·77	4·83	− 1·04 ± ·60	1·73	− 4·76 ± ·75	6·35
38	− ·18 ± ·56	·32	3·09 ± ·66	4·53	2·91 ± ·66	4·41
39	− ·39 ± ·55	·71	·41 ± ·41	1·00	·02 ± ·56	·04
40	− ·64 ± ·69	·93	− 2·34 ± ·51	4·39	− 2·98 ± ·63	4·73

somewhat higher than that of the matured series, C, but the difference in the relative variability may be in part due to the fact that the mean of A is low as compared to that of C.

TABLE XV.

Series Compared	Differences for Ovules per Locule		Differences for Total Ovules per Ovary	
	Positive Differences	Negative Differences	Positive Differences	Negative Differences
All Differences				
B—A	18	10	16	12
C—B	11	17	8	20
C—A	7	21	7	21
Significant Differences				
B—A	10	8	8	7
C—B	6	16	5	17
C—A	5	14	2	15

The results for variability are not so conclusive as those for type, but I think we may safely conclude:

(*a*) That both the eliminated ovaries and the matured fruits have a lower absolute and relative variability than the original population of ovaries.

(*b*) This is due to the ovaries with the lower number of ovules failing to develop, and to those with the higher numbers having the better opportunity to reach maturity.

Problem 3. Is there a difference in the radial asymmetry of the eliminated ovaries and those which develop to maturity?

From the nature of the thin-walled, inflated "bladder" fruit we could not expect it to be other than somewhat irregular in form. The externally visible irregularities in the evaginations of the three locules are of such a nature that no biologist would describe the fruit as asymmetrical. Externally there is no conveniently measurable character from which the amount of irregularity might be measured. The only character which can be determined with certainty is the number of seeds developing or the number of ovules formed per locule.

Opening the three locules of a fruit quite at random—there being no external character to indicate any differentiation whatever between them—we may find such numbers of ovules per locule as:

11—11—11,
10—11—10,
8— 9—11,
9— 7—11,
11— 7—6.

The first of these is clearly enough a radially symmetrical fruit with respect to number of ovules per locule. Any fruit in which the number of ovules is the same in each locule, as 6—6—6, or 7—7—7, is radially symmetrical with respect to this character. Those in which the number of ovules differs from locule to locule are quite as obviously irregular or radially asymmetrical with respect to number of ovules.

Admitting that fruits in which all the locules produce the same number of ovules are radially symmetrical, while those with at least one of the locules differing from the others in its number of ovules are radially asymmetrical, we may be either (*a*) content to divide our ovaries into two classes, radially symmetrical and radially asymmetrical, or (*b*) get some measure of the amount of asymmetry in individual fruits so that the asymmetrical fruits may be subdivided for further analysis.

The measure of asymmetry must be one for the individual fruit, not for a population. The measure must also be independent of the order in which the three locules of the fruit are taken, for there is no constant differentiation between them and they may be opened in any order. Two measures have occurred to me.

First, the sum of the positive difference between the number of ovules in the locules of an ovary might be used. In a fruit of the formula

$$(a)\ (b)\ (c)$$
$$7—8—6,$$

we have, taking all possible differences,

$$a-b=-1, \qquad b-c=+2,$$
$$a-c=+1, \qquad c-a=-1,$$
$$b-a=+1, \qquad c-b=-2,$$

Sum of positive differences $=4$.

Comparing a fruit of the formula 7—8—7 we find the sum of the positive differences = 2. The first is more asymmetrical than the second. This is obvious in the present case from mere inspection.

The second measure is the square root of the mean square deviation of the number of ovules per locule from the mean number in the whole fruit. For the first illustration the mean number per locule is 7 and the deviations are

$$a=0, \quad a^2=0, \qquad \text{Coefficient of asymmetry} = \sqrt{\tfrac{2}{3}} = \cdot 8165.$$
$$b=+1, \quad b^2=1,$$
$$c=-1, \quad c^2=1.$$

For the second illustration: $A = 7 \cdot 3333$, the deviations are

$$a=-\cdot 3333, \quad b=+\cdot 6666, \quad c=-\cdot 3333,$$

and the coefficient of asymmetry is

$$\sqrt{\frac{\cdot 3333^2+\cdot 6666^2+\cdot 3333^2}{3}} = \cdot 4714.$$

This measure is, in short, the standard deviation of the locules of the fruit around their own mean. According to this measure the asymmetry of a fruit with the formula 7—8—7 bears the ratio to one of the formula 7—8—6 of ·471 : ·817 instead of 2 : 4 as indicated by the interlocular difference method.

Since the locules of a fruit are undifferentiated it does not matter how we arrange them for convenience of treatment. I have always copied the results of original countings off in a descending series. The two constants for a few illustrative ovaries are:

Formula	Interlocular Difference	Coefficient of Asymmetry
11—11—11 =	0 =	·0000
11—10—10 =	2 =	·4714
11—10— 9 =	4 =	·8165
11— 9— 9 =	4 =	·9428
11— 9— 8 =	6 =	1·2472
11— 8— 8 =	6 =	1·4142
11— 9— 7 =	8 =	1·6330
11— 8— 7 =	8 =	1·6997
10— 6— 6 =	10 =	1·8856
10— 7— 5 =	10 =	2·0548
11— 7— 6 =	10 =	2·1602

These illustrations show that while the interlocular difference does not distinguish between the amount of irregularity of certain fruits, the coefficient of asymmetry, as I have called the standard deviation of the locules of a fruit around their own mean, does. By mere inspection I am quite unable to decide whether a fruit of the formula 10—6—6 is more irregular, or radially asymmetrical, than one of the formula 10—7—5. Both have the same mean number of ovules per locule and the interlocular difference is the same for both, but the coefficient of asymmetry is slightly higher for the formula 10—7—5.

I have adopted the coefficient of asymmetry in this paper for the following reasons.

(*a*) It is merely the standard deviation—so universally employed in modern statistical work—of the number of ovules per locule in an individual fruit.

(*b*) It differentiates—whether with quantitative accuracy or not—between degrees of asymmetry not distinguished by the interlocular difference method.

(*c*) From data tabled in coefficient of asymmetry classes the interlocular differences may be obtained by the use of a table. The converse is not true.

There is one patent objection to the use of the coefficient of asymmetry as defined here: it is not independent of number of seeds per locule. There are reasons for regarding fruits of the formula 10—6—8 and 5—3—4 as equally asymmetrical, but according to our coefficient their asymmetries bear to each other the ratio of the square roots of 8/3 and 2/3.

The simplest method of freeing our constant from the influence of the absolute number of ovules would seem to be to take the ratio of the coefficient of asymmetry to the mean number of ovules per locule. The reason that I have

not done this is that I have used the coefficient of asymmetry in a large series of investigations on the relationship between asymmetry and fecundity in this and other species*. In them I could not well use the relative coefficient because of the danger of introducing spurious correlation in determining the interdependence of asymmetry and fecundity. It seems desirable, for reasons which will be apparent when my data for fecundity are ready to publish, to use the same method of describing the irregularity of the ovaries in both studies.

The chief objection to the use of the coefficient of asymmetry will be removed if we show that the magnitude is not greatly dependent upon the number of ovules per locule. This must be done for every species investigated. For the present series I have discussed this point in Section VI (see page 493) where it has been shown that the correlation between the coefficient of asymmetry and the number of ovules per locule is very slight.

Table XVI gives the interlocular difference for all the coefficients of asymmetry found in the present study. Anyone who desires may calculate all the constants dealt with in terms of interlocular differences by the use of this table and the tables of data.

TABLE XVI.

Interlocular Differences for Coefficients of Asymmetry.

Coefficient of Asymmetry	Interlocular Difference
·0000	0
·4714	2
·8165	4
·9428	4
1·2472	6
1·4142	6
1·6330	8
1·6997	8
1·8856	10
2·0548	10
2·1602	10

To determine whether there is a selective elimination depending upon the asymmetry of the fruit, the asymmetry of each individual fruit is determined, and the distribution of the asymmetries of the eliminated and uneliminated ovaries compared. The determination of the asymmetries of individual fruits, a rather laborious process at best, is facilitated by tabling the coefficients for all the different ovule formulae at the outset.

A standard deviation calculated on only three "observations"—the three locules of an ovary—is of little value as a means of predicting the standard

* I hope that some portions of these investigations on fecundity will be ready for publication in a few months.

deviation of another fruit of the same species because of the large probable error, but as a description of the irregularity of an individual fruit it is perfectly trustworthy to as many places as we care to carry our arithmetic. I believe the four places to which I have tabled are arithmetically correct. For an individual fruit or very small sample this refinement is quite without significance, but where several hundreds of fruits fall in the same class, as is the case in this work, and especially in large series of data on other problems concerning asymmetry to be published later, the final decimal places may be of significance.

TABLE XVII.

Mean Asymmetry for Individuals.

Shrub	A	B	C
11	·305	·373	·325
12	·488	·456	·429
13	·479	·346	·536
14	·325	·400	·389
15	·478	·470	·424
16	·457	·470	·328
17	·462	·492	·441
18	·397	·435	·270
19	·539	·538	·296
20	·644	·494	·374
21	·401	·267	·339
22	·419	·387	·185
23	·426	·359	·380
24	·491	·454	·407
25	·487	·676	·340
26	·500	·470	·359
27	·459	·514	·332
29	·422	·430	·466
30	·397	·366	·402
31	·475	·436	·400
32	·372	·360	·287
33	·435	·491	·417
34	·599	·549	·362
35	·421	·326	·275
37	·589	·534	·626
38	·285	·361	·339
39	·361	·403	·425
40	·496	·463	·380

TABLE XVIII.

Comparison of Asymmetry.

Shrub	B—A	C—B	C—A
11	·068	−·048	·020
12	−·032	−·027	−·059
13	−·133	·190	·057
14	·075	−·011	·064
15	−·008	−·045	−·053
16	·012	−·142	−·129
17	·030	−·051	−·021
18	·038	−·165	−·127
19	−·001	−·242	−·243
20	−·150	−·120	−·270
21	−·134	−·028	−·162
22	−·032	−·202	−·235
23	−·067	·021	−·046
24	−·036	−·047	−·083
25	·188	−·335	−·147
26	−·030	−·111	−·141
27	·056	−·182	−·126
29	·008	·036	·043
30	−·031	·037	·005
31	−·039	−·036	−·075
32	−·012	−·073	−·085
33	·056	−·074	−·018
34	−·051	−·187	−·238
35	−·095	−·051	−·146
37	·054	·091	−·037
38	·077	−·023	·054
39	·042	·021	·063
40	−·033	−·083	−·116

The frequencies of the different coefficients of radial asymmetry for each of the individual trees are given for the three collections in Tables XXXIV—XXXVI.

The comparison may first be made on the basis of the mean asymmetry of the eliminated and the developing ovaries.

Table XVII gives the mean asymmetry for the three collections from the 28 trees. The excessive labour involved in calculating the probable errors for these 84 means, where the standard deviations must be calculated for classes so

irregular as our asymmetry grades, has led me to omit the probable errors for the individuals. The results for Series A and C are shown graphically in Diagram 4.

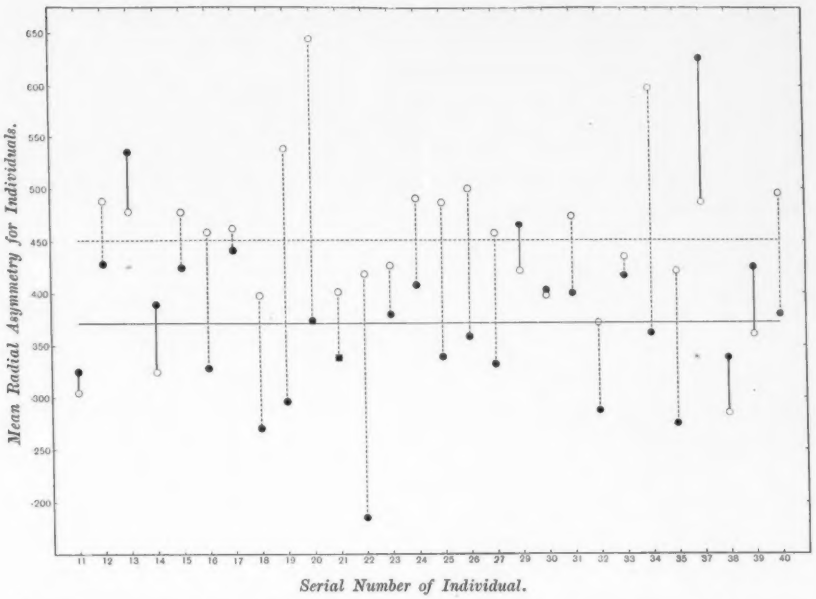


DIAGRAM 4. Showing mean difference in asymmetry of eliminated and matured ovaries for the 28 individuals and for the population. Circles = mean asymmetry of eliminated ovaries for individuals; solid dots = mean asymmetry for matured ovaries for individuals. Broken transverse line = mean asymmetry of all eliminated ovaries; solid transverse line = mean asymmetry of all matured ovaries.

The usual comparisons (B—A, C—B, C—A) are given in Table XVIII. The differences are taken in such a way that if selection tends to eliminate the less symmetrical pods the differences will have the negative sign. For the differences we find:

B—A, 12 positive and 16 negative,

C—B, 6 positive and 22 negative,

C—A, 7 positive and 21 negative.

The mean difference per individual is:

$$B—A = -\cdot0064,$$

$$C—B = -\cdot0675,$$

$$C—A = -\cdot0804.$$

For the grand totals the results are:

Mean asymmetry, Lot A = ·4515 ± ·0051,
Mean asymmetry, Lot B = ·4415 ± ·0048,
Mean asymmetry, Lot C = ·3724 ± ·0045,
B—A = – ·0100 ± ·0070,
C—B = – ·0691 ± ·0066,
C—A = – ·0791 ± ·0068.

The first difference is probably not significant; the second and third are over ten times their probable errors.

The difference in the distributions of asymmetry for the eliminated and the matured series is rather strikingly shown when the frequencies for the two series are reduced to percentages and plotted in a polygon. In Diagram 5 the relatively few asymmetries of 1·2472 and over have been graphed together.

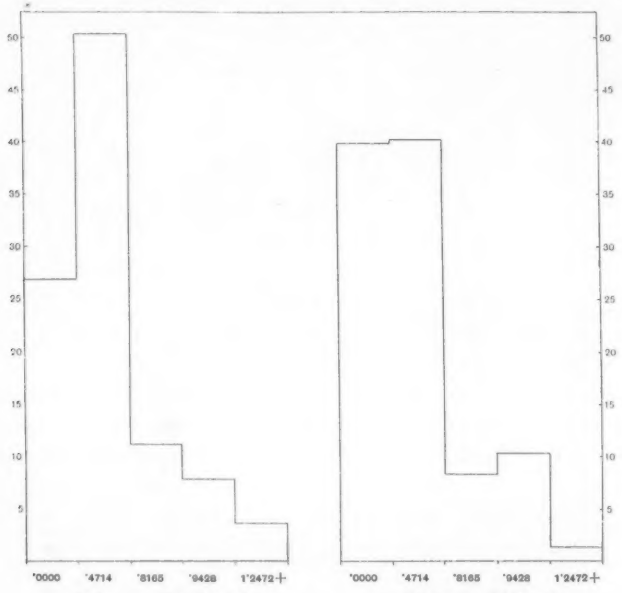


DIAGRAM 5. Percentage frequencies of chief asymmetry classes in the eliminated and matured series.

Whatever way we arrange our data the conclusions flowing from them are the same—that the radial asymmetry of the ovaries which develop to maturity is less than that of those which are eliminated.

Problem 4. Is there a difference in the proportional number of locules with "odd" numbers of ovules in the eliminated ovaries and those which develop to maturity?

Locules with 4, 6, 8, 10 or 12 ovules may be conveniently designated as "even" while those with 5, 7, 9 or 11 ovules may be described as "odd."

With respect to this character the fruits fall into four classes:

3 "even,"
2 "even" + 1 "odd,"
1 "even" + 2 "odd,"
3 "odd."

To the reader it may at first seem to be mere juggling with numbers to seek to determine whether there is any selective elimination of fruits depending upon the number of "odd" or "even" locules per fruit. Nevertheless purely embryological considerations were the incentive to the work.

I have not been able to find any account of the development of the ovary of *Staphylea*, but from the appearance of the matured organ and from our knowledge of the embryology of other forms it would seem quite likely that the ovules of a locule belong to the two margins of the same carpel. If this is the case, locules with "odd" numbers of ovules must have a different number produced on the two margins, while those with an even number almost as surely have the same number on the two carpellary margins. If this be the case no one will deny that structurally "even" locules are more perfectly formed than the "odd," being bilaterally symmetrical with respect to the number of ovules found on the two margins.

The morphology of the fruit can be thoroughly understood only after an embryological investigation, but in the meantime the sensible thing to do seems to be to investigate the question of selective elimination with respect to this point on the basis of the material in hand.

A reason for investigating this point lying within our material itself is the different frequency of fruits with "even" and "odd" locules. For the totals of the three collections the number of fruits belonging to each class is seen in our Table XIX. The predominance of locules with an even number of ovules suggests that this type may represent a point of morphological stability*. If such be the case the determination of whether there is a tendency towards the elimination of those with the larger number of "odd" locules is obviously important.

The test may be made in two ways: (*a*) by the comparison of the percentage of "odd" locules in the three series, and (*b*) by calculating the percentage of each of the four different ovule formulae in the three collections.

* Compare Sir Francis Galton's suggestion concerning points of organic stability in natural inheritance.

The percentage of total locules with "odd" numbers of ovules in the three series is:

Percentage "odd" in Lot A = 37·407 ± ·412,
Percentage "odd" in Lot B = 36·065 ± ·377,
Percentage "odd" in Lot C = 25·185 ± ·325.

The probable errors for these frequencies were calculated from the formula

$$Ey_s = \cdot 67449 \sqrt{y_s \times \left(1 - \frac{y_s}{m}\right)},$$

where y_s is the frequency in any one of the four classes, and m is the total population. From the probable error for the absolute frequencies so obtained the probable errors for the percentage frequencies were got by taking the ratio Ey_s/m.

Comparing the series among themselves as usual we get:

B—A = − 1·342 ± ·558 %,
C—B = − 10·880 ± ·498 %,
C—A = − 12·222 ± ·524 %.

In all these cases there is an actual decrease in the relative number of "odd" locules as elimination progresses. The decrease of 1·3 per cent. noted from A to B is perhaps not significant, being only about twice its probable error, but the fall of 11 per cent. from B to C is about 22 times its probable error and the 12 per cent. decrease from A to C is over 23 times its probable error.

There can be no reasonable doubt, therefore, that "odd" are more likely to be eliminated than "even" locules. Naturally one locule cannot be eliminated alone, but it has seemed interesting to work through the results for individual locules. We turn now to the more difficult, but more significant, comparison of individual formulae.

The frequencies, absolute and relative, appear in Table XIX. Diagram 6 shows

TABLE XIX. *Frequency of Ovaries of Different Locular Composition in the three Series.*

Formula	A		B		C	
	f	%	*f*	%	*f*	%
3 "even" ...	595	28·401	741	30·061	1265	46·783
2 "even," 1 "odd"	788	37·613	936	37·970	928	34·320
1 "even," 2 "odd"	573	27·351	633	25·680	418	15·458
3 "odd"	139	6·635	155	6·288	93	3·439
Totals	2095	100·000	2465	99·999	2704	100·000
Even Locules ...	3934	62·593	4728	63·935	6069	74·815
Odd Locules ...	2351	37·407	2667	36·065	2043	25·185

the results graphically. The three comparisons are made on the basis of the percentage frequencies in Table XX.

TABLE XX.

Formula	B—A	C—B	C—A
3 "even" ...	+1·660	+16·722	+18·382
2 "even," 1 "odd"	+ ·357	− 3·650	− 3·293
1 "even," 2 "odd"	−1·671	−10·222	−11·893
3 "odd"	− ·347	− 2·849	− 3·196

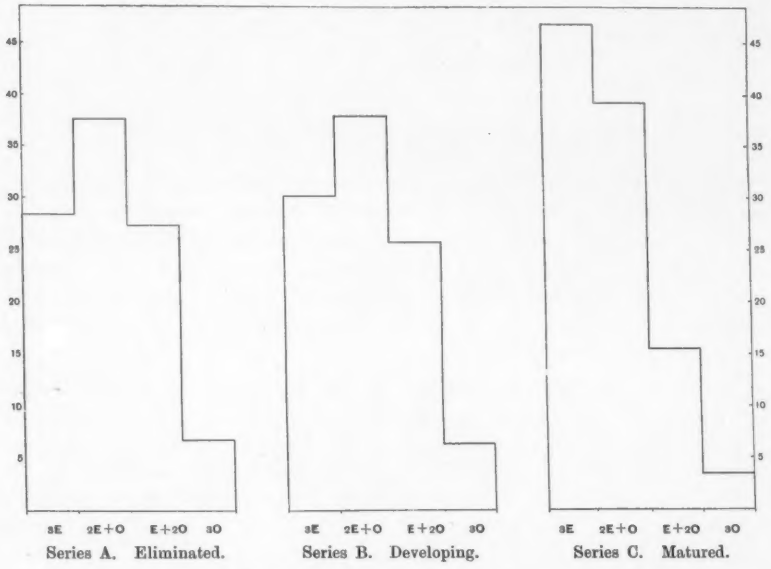


DIAGRAM 6. Percentage frequencies of four types of locular composition of ovaries in the three series.

In both of the critical comparisons it appears that elimination brings about an increase in the per cent. of ovaries with all three locules "even" and a decrease in the per cent. of those which contain one or more locules with an odd number of ovules. The differences are also conspicuously large, amounting to 18·4 per cent. in the case of the "all even" class.

In view of the results secured for the individual locules, it has not seemed necessary to calculate the probable errors of their differences. A better test would seem to be that for the probability against a given system of deviations from theory, in the case of a correlated system of variables having arisen through the

errors of random sampling*. In making this test we calculate the frequencies to be expected in the classes of a series n if they occurred in the same proportionate frequency as in the series m, and determine the chances against this being due merely to the errors of random sampling from Elderton's Tables†.

Only the critical comparison C—A need be tested. Table XXI gives the data. χ^2 equals 510·8, a value over seven times as large as Elderton has thought it

TABLE XXI.

Formula	Observed Frequency	Calculated Frequency	$m_r - m_r'$	$\frac{(m_r - m_r')^2}{m_r}$
3 "even" ...	1265	768	−497	321·6
2 "even," 1 "odd"	928	1017	+ 89	7·8
1 "even," 2 "odd"	418	740	+322	140·1
3 "odd"	93	179	+ 86	41·3
Totals ...	2704	2704	0	510·8

worth while to table. The chances are thus many millions to one against such a divergence as this occurring by the accidents of random sampling.

Why locules with an odd number of ovules should be inferior to those with an even number, or fruits composed in part or altogether of "odd" locules should be inferior to those composed chiefly or altogether of "even" locules, our data do not permit us to determine. There can be no reasonable doubt of the fact of the relationship for our material.

Problem 5. Is there a difference in the percentage of 2- and 4-merous ovaries among the eliminated ovaries and those which develop to maturity?

The answer to this question cannot be expected to be very conclusive because of the rarity of the anomalies in the series under consideration. The available data are given in Table XXXVII. The reason for the frequencies for 3-merous fruits being higher in Series C in this Table than in those given before is that, to determine more accurately the percentage of anomalous fruits, larger samples were examined for this point without counting the numbers of ovules.

For present purposes using only the grand totals, we find the frequencies for 2-loculed fruits:

Series A = 2·477 per cent.
Series B = 2·321 per cent.
Series C = 1·488 per cent.
B—A = − ·156 per cent.
C—B = − ·833 per cent.
C—A = − ·989 per cent.

* Pearson, K.: *Phil. Mag.* Vol. L. pp. 157—175, 1900.
† Elderton, W. P.: *Biometrika*, Vol. I. pp. 155—163, 1901.

For 4-merous fruits:

Series A = 1·376 per cent.
Series B = 2·243 per cent.
Series C = 2·461 per cent.
B—A = + ·867 per cent.
C—B = + ·218 per cent.
C—A = + 1·085 per cent.

No probable errors have been calculated for the reason that our percentages are so low that the familiar theory of probable errors cannot be applied.

It is quite impossible to say, therefore, whether or no these differences are really significant, but so far as our data go they indicate that the 2-merous ovaries are more likely to be eliminated than the 3-merous, while the 4-merous ones are less likely to be eliminated.

V. Discussion of Problems on the Basis of Actual Data for 1906.

Problem 1. The seriations of the number of ovules per locule are found in Table XXXVIII; those for total ovules per fruit in Table XXXIX.

For constants we get:

Series A (6—10 mm.), Mean = 7·232 ± ·029.
Series B (11—15 mm.), Mean = 7·660 ± ·014.
Series C (16—20 mm.), Mean = 7·821 ± ·016.
B—A = + ·428 ± ·032.
C—B = + ·161 ± ·021.
C—A = + ·589 ± ·033.

The increase from A to B is 13·4 times its probable error, that from B to C is 8·6 times its probable error and the difference between the mean of the smallest and the largest fruits is 17·8 times its probable error. The increase from the smallest to the largest fruits is 8·14 per cent. of the value for the former.

These results certainly substantiate in the most unequivocal way those secured for means from the 1908 series, and are the more interesting from the fact that the material was collected not merely for a different purpose and with no idea of testing selective elimination, but by a quite different method.

Problem 2. The absolute and relative variabilities for both ovules per locule and total ovules per ovary may again be considered. The complete answer to this problem is given in Table XXII.

As the fruits become larger the variability becomes less. For the absolute variabilities the probable errors show that there is a very high probability that every difference is significant. Probable errors have not been calculated for the coefficients of variation, but the differences are all relatively large and consistent.

TABLE XXII.

Comparison of Absolute and Relative Variabilities for 1906.

Material	Standard Deviation and Probable Error	Coefficient of Variation
Ovules per Locule		
Series A ...	·968 ± ·020	13·382
Series B ...	·774 ± ·010	10·104
Series C ...	·685 ± ·011	8·762
B—A ...	− ·194 ± ·022	− 3·278
C—B ...	− ·089 ± ·015	− 1·342
C—A ...	− ·283 ± ·023	− 4·620
Ovules per Ovary		
Series A ...	2·497 ± ·090	11·528
Series B ...	1·895 ± ·042	8·247
Series C ...	1·594 ± ·046	6·795
B—A ...	− ·602 ± ·099	− 3·281
C—B ...	− ·301 ± ·062	− 1·272
C—A ...	− ·903 ± ·101	− 4·553

Problem 3. The frequency of the different asymmetry classes in the three series is given in Table XL. The three differences of the relative frequencies are given in Table XXIII.

TABLE XXIII.

	B—A	C—B	C—A
·0000	+3·751	+8·495	12·246
·4714	+3·064	−9·790	− 6·726
·8165	−4·877	+2·067	− 2·810
·9428	−1·790	−1·077	− 2·867
1·2472	—	—	—

The results show that as the size of the fruit increases the symmetrical fruits form a higher percentage of the entire population, while the asymmetrical fruits of the various grades become relatively fewer. Our numbers are small and some irregularity is to be expected.

The average asymmetries are:

A, 6—10 mm. long, Mean = ·3822 ± ·0160.
B, 11—15 mm. long, Mean = ·3381 ± ·0091.
C, 16—20 mm. long, Mean = ·3025 ± ·0126.

B—A = − ·044 ± ·018.
C—B = − ·036 ± ·016.
C—A = − ·080 ± ·020.

There is clearly a decrease in asymmetry from A to B and again from B to C. Owing to the smallness of our series the probable errors are high, but all the differences are over twice their probable errors, and the critical difference C—A is four times its probable error.

Diagram 7 shows graphically the frequency of the four chief asymmetry classes.

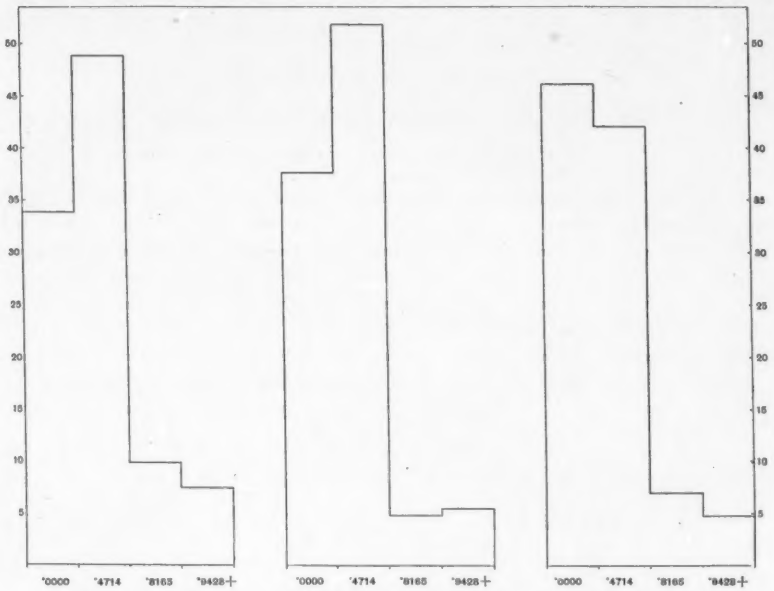


DIAGRAM 7. Percentage frequencies of chief asymmetry classes in the three series of developing fruits for 1906.

Problem 4. The frequency of the four types of ovaries with respect to the number of "odd" locules produced is shown for the three lengths of fruit in Table XXIV. and graphically (with percentage frequencies) in Diagram 8.

TABLE XXIV. *Frequency of Ovaries of Different Locular Composition in the three Series.*

Formula	6—10 mm. =A		11—15 mm. =B		16—20 mm. =C	
	f	%	*f*	%	*f*	%
3 "even" ...	58	33·333	169	35·957	125	45·787
2 "even," 1 "odd"	68	39·080	162	34·468	85	31·136
1 "even," 2 "odd"	35	20·115	107	22·766	51	18·681
3 "odd"	13	7·471	32	6·809	12	3·396

The reader will note, by comparing the percentage frequencies for the smallest and the largest fruits, that the 3 "even" class is much more abundantly represented in the largest than in the smallest fruits. Those with one or more "odd" locules are necessarily less abundant.

The percentage frequency of "odd" locules in the three series is:

Lot A, 6—10 mm. = 33·91 ± 1·40 °/₀.

Lot B, 11—15 mm. = 33·48 ± ·85 °/₀.

Lot C, 16—20 mm. = 27·23 ± 1·05 °/₀.

The difference, C—A = − 6·68 ± 1·75 °/₀.

In predicting the frequencies of the four classes in Series C from the distribution found in Series A, to ascertain whether the former could possibly be regarded as arising from the latter through the errors of random sampling, we find

$$\chi^2 = 20{\cdot}7,\ P = {\cdot}00012.$$

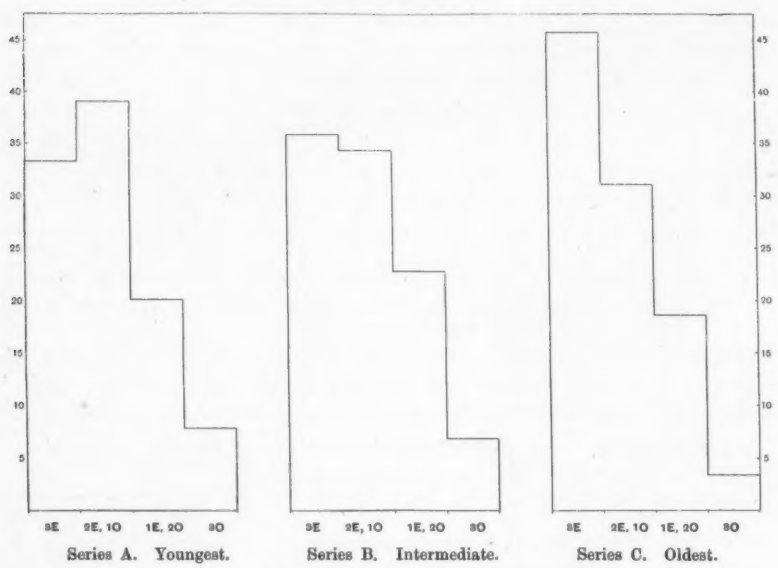


DIAGRAM 8. Percentage frequencies of four types of locular composition of ovaries in the three series of developing fruits for 1906.

Thus deviations of observation from theory so great as those found here would be expected to arise through the errors of random sampling only about 12 times out of 100,000.

Allowing full weight to the fact that the frequencies in series A have their probable errors, I think we can maintain that series C and series A are differentiated from each other with respect to the number of "odd" and "even" locules in their ovaries.

Problem 5. This problem cannot be discussed, since no data were collected for other than 3-merous fruits.

VI. INTERRELATIONSHIP OF CHARACTERS CONSIDERED.

For convenience and clearness our characters have been treated in the foregoing discussion of problems as though they were quite independent.

Analysis can be carried somewhat further. If a character is quite uncorrelated with others, we shall be confident in regarding any selective elimination which we find associated with this character as arising through some unfitness for continued development associated with it. If, on the contrary, it appears that a character is correlated with some other, we cannot know without further evidence which of a pair is potent in producing a selective elimination.

These points will best be made clear by an examination of the three relationships which seem worth consideration for our material.

1. *The Correlation between the Coefficient of Asymmetry and the Number of "odd" Locules per Ovary.*

The fact that both radial asymmetry as measured by the deviation of the number of ovules per locule from their mean and the composition of the ovary with regard to the "even" and "odd" nature of the individual locules seem to be of significance in determining whether it shall develop into a mature fruit, does not necessarily prove that with *both* of these characteristics there is associated some functional unfitness for development.

There is necessarily *some* correlation between the two characteristics in any fruit. If the three locules have all the same number of ovules and they be even, say 8—8—8, the asymmetry will be ·0000. If it be "2 even, 1 odd," say 8—8—7 or 9—8—8, the asymmetry will be ·4714. Thus a fruit composed of "odd" and "even" must necessarily be somewhat asymmetrical while one of "all even" or "all odd" may be perfectly asymmetrical. If an "all even" or "all odd" fruit is radially asymmetrical with respect to number of ovules per locule, it may be more asymmetrical than a fruit of locules of both types, since there must be a difference of two ovules between two different "even" locules, while an "even" and an "odd" may differ by only a single ovule. For instance, 8—8—6 is more asymmetrical than 8—8—7.

As the reader will see, the problems are very complex. Even the relatively large series of data available for this research are, I fear, insufficient for a detailed study of the questions of the interdependence of these two characteristics of the fruit and their fitness. Later I hope to publish extensive data already on hand that may throw some light on these questions. For the present I shall only show the correlation between the coefficient of asymmetry and the number of "odd" locules per fruit. The data for the three series appear in Tables XLI—XLIII, and the constants by the product moment method* are given in Table XXV.

TABLE XXV.

	Correlation and Probable Error
Series A ...	·2247 ± ·0140
Series B ...	·3296 ± ·0121
Series C ...	·3282 ± ·0116

Clearly there is a definitely significant, though not very large, interdependence between the two characteristics.

It may be possible, therefore, that one of these characteristics has comparatively little independent significance in determining the fitness of an ovary for continuing its development. If only one of the two characteristics is thus potent while the influence of the second is only apparent because of its dependence upon the first, we cannot determine at present which is independent and which is dependent.

Personally, I believe that the coefficients of correlation between them are too low to account entirely for the results for elimination that have been secured, and that therefore they are both concerned in determining whether or no any ovary shall develop. My reasons are as follows:

If the elimination of the ovaries of one characteristic, a, of the two here considered be not primarily due to their possessing their first characteristic but merely apparently connected with it because it is itself correlated with the second characteristic, b†, we should expect no selective elimination with respect to the dependent character within the subgroups of the independent character. Concretely, if the elimination of ovaries with a larger number of "odd" locules is due solely to the fact that these ovaries are also more radially asymmetrical, we would expect to find no elimination with regard to number of "odd" locules when we work within the same asymmetry class. Similarly for elimination with respect to radial asymmetry. Tersely: for constant a is there an elimination depending on b; for constant b is there an elimination depending on a?

* Calculated without any combination or modification of classes for the coefficients of asymmetry.

† Pearson, K.: *Phil. Trans.* A., Vol. cc. pp. 18, 19, 1902, has termed this indirect selection.

Data are available in Tables XLI—XLIII. Only the critical comparison C—A need be made. Table XXVI. gives the difference for the mean number

TABLE XXVI.

Radial Asymmetry	Mean "Odd" Locules, C—A
·0000	−·335
·4714	−·106
·8165	−·115
·9428	−·454
1·2472	−·157

of "odd" locules for the first five radial asymmetry classes. Beyond this the frequencies are too few to be trustworthy. Table XXVII. gives the difference

TABLE XXVII.

Character of Ovary	Mean Radial Asymmetry, C—A
3 "even" ...	−·017
2 "even," 1 "odd"	−·019
1 "even," 2 "odd"	−·034
3 "odd"	−·077

between the mean radial asymmetry of C and A for the four classes of ovaries with respect to number of "odd" locules per ovary.

I attach no importance to the numerical value of these differences, for I have not calculated their probable errors, but the negative sign throughout seems to me fairly satisfactory evidence that both of these characters are to some extent of independent significance in determining whether an ovary shall develop to maturity.

2. *The Correlation between the Coefficient of Asymmetry and the Number of Ovules per Locule.*

Both the actual number of ovules per locule—or per ovary—and the radial asymmetry of the fruit with respect to the number of ovules per locule seem to have significance in determining the fate of an ovary. Are they really both significant, or is one dependent upon the other?

This is a fundamental physiological question. I believe a trustworthy answer is given by the coefficient of correlation between the asymmetry of the fruit and the number of ovules per locule. Series B is the most like the original popula-

tion of ovaries and serves for the determination. The coefficient calculated from the surface given as Table XLIV is

$$r = \cdot 0221 \pm \cdot 0078.$$

Possibly the coefficient is statistically significant, being about thrice its probable error, but it is clearly of so low an order that it can have no appreciable biological significance in the problem now in hand.

Apparently, therefore, both number of ovules and radial asymmetry in the distribution of the ovules in the three locules are independent and physiologically significant factors in determining whether an ovary shall or shall not reach maturity.

3. *The Correlation between the Number of "odd" Locules per Ovary and the Number of Ovules per Locule.*

Again, series B may serve as material. The correlation surface is shown as Table XLV. The constant is

$$r = -\cdot 0368 \pm \cdot 0078.$$

With a correlation so low as this it seems reasonable to think that, in their relation to selective elimination, these two characters are practically independent.

VII. Recapitulation.

1. The foregoing pages contain an account of a quantitative study of the selective elimination occurring in the ovaries of *Staphylea trifolia* between the time of flowering and the maturing of the fruit. The scope of the paper is limited strictly to the presentation of observed data and their description by means of biometric constants. No biological theories to explain why one ovary should be less fit than another are suggested, and no hypotheses concerning the influence of the selective elimination demonstrated upon evolutionary progress are brought forward; there will be ample time for this after the much more extensive data already collected are published. A comparison of these results for the selective elimination of *organs* with those secured by others for a selective elimination of *individuals* may also be profitably postponed.

2. The ovaries with relatively low numbers of ovules are more extensively eliminated than those with high numbers. The mean of the population remaining after selection is about 7 or 8 per cent. higher than that of the eliminated individuals.

3. The variability of both the eliminated ovaries and the series remaining after elimination is less than that of the original population. This condition

would be expected as the result of the elimination of the ovaries with the generally lower number of ovules, and the continued development of those with the generally higher numbers, as described immediately above.

4. The ovaries which remain after elimination are more radially symmetrical (as measured by the coefficient of asymmetry, or the standard deviation of the number of ovules per locule around their own mean) than those which are eliminated.

5. Ovaries with one or more locules with an "odd" number of ovules are more likely to be eliminated than those with all the locules bearing an "even" number.

6. Dimerous ovaries seem less likely, and tetramerous ovaries more likely to develop to maturity than the normal trimerous ones. Further data bearing on this point are needed.

7. Radial asymmetry and the composition of the fruit with respect to number of "odd" and "even" locules are not independent, but correlated characters. Evidence is adduced to show that both are to some extent of independent significance in determining whether or no an ovary shall continue its development.

8. Neither radial asymmetry nor the composition of the ovary with respect to number of "odd" and "even" numbers of ovules per locule is closely enough correlated with number of ovules per locule to modify in any way our conclusions concerning them as independent characteristics in their relation to selective elimination.

BIOMETRIC LABORATORY,
UNIVERSITY COLLEGE, LONDON.
January 31, 1910.

VIII. APPENDIX OF TABLES OF DATA.

In publishing these data and the constants calculated from them I have retained for the individuals the original numbers of my notes, 11—40, since I shall later treat other features of the same individuals and it may be a convenience to the reader who cares to make comparisons, as well as to myself.

In titling the Tables the following symbols were used:

L = Number of ovules per locule. O = Total number of ovules per ovary. N = Number of locules per ovary. LC = Locular composition, i.e., the composition of the fruit with respect to number of locules with "odd" and "even" numbers of ovules. A = Asymmetry, as measured by the coefficient of asymmetry.

A, B and C, following the year of collection, refer to the categories explained on p. 456.

TABLE XXVIII—L 1908, A.

Serial Number of Shrub.

Ovules per Locule.	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*37*	*38*	*39*	*40*	Totals
3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	2
4	—	—	—	1	—	—	—	—	—	1	—	—	2	—	—	—	—	—	4	—	—	—	—	—	—	4	—	—	12
5	—	6	1	3	2	10	—	2	14	30	3	1	32	8	2	6	10	4	46	3	—	18	7	4	5	31	—	5	253
6	1	68	14	90	12	59	20	11	44	80	16	13	79	34	23	39	39	144	161	51	15	109	40	28	26	217	4	47	1484
7	24	89	35	44	65	74	70	54	103	73	44	50	29	68	28	61	77	104	24	123	42	93	49	90	43	41	19	67	1683
8	221	123	82	9	169	114	152	81	97	51	84	93	8	67	64	44	66	45	5	118	122	50	46	173	78	5	78	72	2317
9	47	14	16	3	41	13	45	2	12	5	3	8	—	3	46	—	3	3	—	5	16	—	8	5	88	—	19	4	409
10	7	—	2	—	11	—	12	—	—	—	—	—	—	—	32	—	—	—	—	—	—	—	—	—	57	—	—	—	121
11	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3	—	—	—	4
Totals	300	300	150	150	300	270	300	150	270	240	150	165	150	180	195	150	195	300	240	300	195	270	150	300	300	300	120	195	6285

TABLE XXIX—L 1908, B.

Serial Number of Shrub.

Ovules per Locule.	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*37*	*38*	*39*	*40*	Totals
2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
4	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6	1	—	3	4	—	—	9	—	—	24
5	—	15	—	7	—	13	—	10	10	6	—	—	23	13	2	6	6	32	30	10	1	27	13	1	1	63	—	1	290
6	2	104	7	75	5	67	5	78	44	69	12	8	90	43	8	33	38	160	160	86	25	130	60	6	4	203	4	21	1547
7	29	75	64	50	47	78	50	94	82	98	46	67	30	68	24	53	65	76	53	142	65	67	81	74	16	19	21	75	1709
8	202	99	207	15	180	96	165	114	88	84	185	111	7	68	73	57	110	26	4	55	188	27	103	211	109	6	186	160	2936
9	54	6	18	—	57	14	63	4	13	13	12	9	—	3	92	1	19	5	—	6	21	1	8	8	98	—	78	32	635
10	13	—	4	2	7	1	13	—	3	—	—	—	—	—	86	—	2	1	—	—	—	—	1	—	61	—	10	11	215
11	—	—	—	1	4	1	4	—	—	—	—	—	—	—	13	—	—	—	—	—	—	—	—	—	8	—	1	—	32
12	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	3	—	—	—	5
Totals	300	300	300	150	300	270	300	300	240	270	255	195	150	195	300	150	240	300	255	300	300	255	270	300	300	300	300	300	7395

TABLE XXX—L 1908, C.

TABLE XXX—L 1908, C.

Serial Number of Shrub.

Ovules per Locule	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	29	30	31	32	33	34	35	37	38	39	40	Totals
4	—	—	—	—	—	—	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3
5	—	—	2	5	—	1	—	1	1	—	—	—	16	—	—	—	—	4	10	3	2	2	2	1	—	7	—	—	57
6	2	52	165	101	—	12	2	38	17	24	9	9	121	39	3	17	8	98	185	27	16	115	35	12	6	100	5	2	1220
7	14	57	37	71	6	16	12	52	40	75	15	26	24	61	1	59	15	117	84	108	38	114	37	36	31	80	14	26	1266
8	190	186	96	27	139	201	121	203	210	186	236	252	18	200	28	147	233	81	21	153	227	69	219	241	125	111	193	191	4304
9	62	5	—	—	93	40	80	5	32	14	27	11	1	—	59	4	26	—	—	8	15	—	7	8	77	2	65	62	703
10	32	—	—	—	59	30	79	1	—	1	13	2	—	—	200	1	18	—	—	1	2	—	—	2	58	—	23	19	541
11	—	—	—	—	3	—	3	—	—	—	—	—	—	—	9	—	—	—	—	—	—	—	—	—	2	—	—	—	17
12	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
Totals	300	300	300	204	300	300	300	300	300	300	300	300	180	300	300	228	300	300	300	300	300	300	300	300	300	300	300	300	8112

TABLE XXXI—O 1908, A.

Serial Number of Shrub.

Total Ovules per Ovary.	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	29	30	31	32	33	34	35	37	38	39	40	Totals
12	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
13	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
14	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	3
15	—	—	—	1	—	—	—	—	—	—	—	—	5	—	—	—	1	—	3	—	—	1	—	—	—	1	—	—	12
16	—	—	—	1	—	2	—	—	1	3	—	—	2	—	—	1	1	1	8	—	—	1	—	—	1	6	—	—	28
17	—	1	—	—	—	3	—	—	3	6	1	1	11	2	2	—	2	1	21	1	—	7	—	—	1	17	—	2	82
18	—	9	—	18	1	8	2	1	3	11	2	—	8	3	1	6	3	21	30	2	2	17	8	1	2	49	—	4	212
19	—	4	2	13	—	4	1	1	6	10	1	1	10	2	4	5	6	24	11	12	—	10	5	5	3	11	—	8	159
20	—	18	2	6	—	7	4	3	13	15	4	2	8	15	4	7	8	18	6	8	3	16	6	6	5	8	—	6	198
21	1	8	5	3	12	14	6	5	12	13	4	5	3	9	1	11	12	12	1	14	2	18	8	11	4	4	4	13	215
22	—	23	9	4	9	11	14	16	25	13	7	14	2	10	4	10	14	11	—	28	8	12	8	20	6	1	4	10	293
23	18	13	9	3	20	18	21	10	11	5	17	13	—	10	6	7	9	10	—	21	19	5	6	26	6	1	6	12	302
24	46	19	15	1	32	19	20	14	13	3	13	14	—	7	11	3	7	2	—	13	25	3	8	30	9	—	13	9	349
25	21	3	6	—	12	2	18	—	1	—	1	5	—	2	6	—	2	—	—	1	4	—	1	1	14	—	9	1	110
26	9	2	—	—	8	2	6	—	2	—	—	—	—	—	6	—	—	—	—	—	1	—	—	—	12	—	3	—	51
27	3	—	1	—	4	—	3	—	—	—	—	—	—	—	4	—	—	—	—	—	1	—	—	—	12	—	1	—	29
28	2	—	—	—	2	—	4	—	—	—	—	—	—	—	8	—	—	—	—	—	—	—	—	—	12	—	—	—	28
29	—	—	1	—	—	—	—	—	—	—	—	—	—	—	7	—	—	—	—	—	—	—	—	—	8	—	—	—	16
30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	4	—	—	—	5
31	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	2
Totals	100	100	50	50	100	90	100	50	90	80	50	55	50	60	65	50	65	100	80	100	65	90	50	100	100	100	40	65	2095

TABLE XXXII—O 1908, B.

Serial Number of Shrub.

Total Ovules per Ovary.

	11	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*37*	*38*	*39*	*40*	Totals
11	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1
12	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
13	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	1	—	—	2
14	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—	—	1	—	—	3
15	—	1	—	—	—	—	—	—	1	—	—	—	2	—	—	—	—	1	2	—	—	—	2	—	—	4	—	—	13
16	—	3	—	1	—	1	—	2	1	—	—	—	3	2	—	—	1	7	3	2	—	5	1	—	—	13	—	—	45
17	—	3	—	3	—	7	—	2	1	1	—	—	5	4	—	1	—	8	10	6	—	8	4	—	—	29	—	—	92
18	—	13	—	11	—	6	—	10	3	8	—	—	21	5	1	3	6	32	39	4	3	21	5	—	—	35	1	1	228
19	—	15	—	12	—	8	1	9	6	7	3	1	9	4	1	7	2	16	13	12	3	20	8	1	1	13	—	1	173
20	—	12	1	8	—	14	1	16	12	18	1	2	4	8	—	6	4	9	9	28	5	13	9	1	—	2	—	7	190
21	—	15	5	9	3	6	1	10	11	17	5	6	4	10	2	10	14	10	5	21	2	9	16	3	1	1	—	7	203
22	1	12	11	3	6	14	6	14	14	11	5	12	1	15	2	7	10	11	1	8	17	3	12	16	3	—	2	16	233
23	14	7	23	1	20	12	20	19	9	12	24	22	1	6	6	9	15	4	—	13	21	2	17	36	6	—	10	23	352
24	52	15	52	1	32	14	30	18	13	13	40	20	—	9	10	6	21	2	—	3	35	3	13	39	18	1	34	18	512
25	18	3	5	—	21	4	20	—	9	1	5	1	—	2	8	1	4	—	—	2	11	—	3	3	8	—	26	13	168
26	8	1	2	—	13	3	8	—	—	1	1	1	—	—	12	—	2	—	—	1	3	—	—	1	14	—	18	10	99
27	1	—	1	—	2	—	6	—	—	1	1	—	—	—	14	—	1	—	—	—	—	—	—	—	15	—	7	2	51
28	4	—	—	—	1	—	4	—	—	—	—	—	—	—	19	—	—	—	—	—	—	—	—	—	14	—	1	1	44
29	2	—	—	—	2	1	3	—	—	—	—	—	—	—	11	—	—	—	—	—	—	—	—	—	12	—	1	—	32
30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	—	—	—	—	—	—	—	—	—	6	—	—	1	19
31	—	—	—	1	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	—	—	—	—	1	—	—	—	4
32	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
33	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
Totals	100	100	100	50	100	90	100	100	80	90	85	65	50	65	100	50	80	100	85	100	100	85	90	100	100	100	100	100	2465

TABLE XXXIII—O 1908, C.

Serial Number of Shrub.

Total Ovules per Ovary.	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	29	30	31	32	33	34	35	37	38	39	40	Totals
12	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
13	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
14	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
15	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
16	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	1	—	—	—	—	—	—	—	1	—	—	4
17	—	—	1	2	—	—	—	—	—	—	—	—	6	—	—	—	—	1	6	—	1	—	—	—	—	4	—	—	21
18	—	2	23	16	—	1	—	6	1	—	—	1	30	2	1	—	—	10	31	4	1	15	2	—	—	15	—	—	161
19	—	5	13	17	—	1	—	4	5	5	—	1	5	1	—	1	—	11	26	2	—	19	—	—	—	12	1	—	129
20	—	8	27	10	—	1	—	4	1	5	2	2	7	5	—	6	1	24	21	7	3	19	8	2	1	11	—	1	176
21	—	10	4	15	1	1	2	5	2	7	1	1	4	10	—	6	1	16	10	15	4	19	8	3	1	12	2	1	161
22	4	21	18	3	—	7	—	13	9	12	6	2	5	22	—	11	5	26	6	16	10	9	14	14	3	12	2	3	247
23	5	17	7	2	3	9	5	16	15	25	8	15	—	28	—	24	7	11	—	28	16	14	16	17	7	10	5	10	320
24	41	35	7	3	16	42	19	48	44	39	61	68	1	32	1	24	56	6	—	23	53	5	46	57	23	22	31	37	840
25	21	2	—	—	25	14	17	3	18	6	10	8	—	—	6	3	15	—	—	4	8	—	6	4	14	1	30	22	237
26	9	—	—	—	21	9	9	1	4	1	4	2	—	—	4	—	9	—	—	1	4	—	—	2	21	—	16	14	131
27	9	—	—	—	8	5	12	—	1	—	3	—	—	—	5	1	4	—	—	—	—	—	—	1	14	—	9	7	79
28	4	—	—	—	8	3	16	—	—	—	2	—	—	—	10	—	1	—	—	—	—	—	—	—	5	—	2	4	55
29	3	—	—	—	12	3	8	—	—	—	2	—	—	—	30	—	—	—	—	—	—	—	—	—	5	—	2	—	65
30	4	—	—	—	5	4	10	—	—	—	1	—	—	—	36	—	1	—	—	—	—	—	—	—	5	—	—	1	67
31	—	—	—	—	1	—	1	—	—	—	—	—	—	—	6	—	—	—	—	—	—	—	—	—	1	—	—	—	9
32	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	100	100	100	68	100	100	100	100	100	100	100	100	60	100	100	76	100	100	100	100	100	100	100	100	100	100	100	100	2704

TABLE XXXIV—A 1908, A.

Serial Number of Shrub.

Coefficient of Asymmetry.	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*37*	*38*	*39*	*40*	Totals
·0000	44	27	13	21	29	25	21	17	17	10	15	14	12	15	13	9	14	26	27	17	25	28	8	31	13	48	13	13	565
·4714	47	44	23	23	41	41	61	25	46	36	28	33	31	30	38	26	37	56	38	62	28	40	19	49	56	43	22	34	1057
·8165	6	8	8	2	17	16	9	2	9	13	4	4	3	4	4	11	9	9	6	12	5	10	15	10	14	6	5	12	233
·9428	1	16	3	4	7	4	5	4	13	11	2	3	3	7	7	4	4	9	8	9	6	10	6	6	7	2	—	4	165
1·2472	2	4	3	—	3	4	3	1	2	5	—	—	—	3	2	—	1	—	—	—	1	1	—	3	4	—	—	1	43
1·4142	—	1	—	—	3	—	—	1	1	4	—	1	1	—	1	—	—	—	1	—	—	1	1	1	3	1	—	1	22
1·6330	—	—	—	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	3
1·6997	—	—	—	—	—	—	—	—	1	1	1	—	—	1	—	—	—	—	—	—	—	—	1	—	1	—	—	—	6
1·8856	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
Totals	100	100	50	50	100	90	100	50	90	80	50	55	50	60	65	50	65	100	80	100	65	90	50	100	100	100	40	65	2095

TABLE XXXV—A 1908, B.

Serial Number of Shrub.

Coefficient of Asymmetry.	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*37*	*38*	*39*	*40*	Totals
·0000	40	28	45	15	26	20	21	28	14	23	43	20	20	18	15	14	21	31	35	21	34	19	21	38	22	35	30	23	720
·4714	40	45	36	26	53	50	55	51	43	38	35	35	20	31	38	24	30	47	34	65	54	41	38	55	44	53	53	56	1190
·8165	12	16	13	6	10	6	13	10	13	16	3	6	7	5	21	4	18	14	10	7	6	14	13	3	19	6	11	6	288
·9428	5	8	3	3	5	12	5	8	3	9	4	4	3	10	10	4	7	4	4	5	6	9	9	3	8	4	5	12	172
1·2472	2	3	2	—	3	2	3	3	5	4	—	—	—	—	9	3	1	3	—	2	—	2	6	—	5	2	—	3	63
1·4142	1	—	—	—	—	—	3	—	1	—	—	—	—	1	3	1	2	—	1	—	—	—	2	1	—	—	—	—	16
1·6330	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1	—	3
1·6997	—	—	1	—	1	—	—	—	1	—	—	—	—	—	2	—	—	—	1	—	—	—	—	—	2	—	—	—	8
1·8856	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1
2·0548	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1	—	—	—	—	—	3
2·1602	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1
Totals	100	100	100	50	100	90	100	100	80	90	85	65	50	65	100	50	80	100	85	100	100	85	90	100	100	100	100	100	2465

TABLE XXXVI—A 1908, C.

Serial Number of Shrub.

Coefficient of Asymmetry.	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*37*	*38*	*39*	*40*	Totals
·0000	46	37	29	23	24	48	30	56	45	35	61	67	28	38	39	27	54	20	31	29	54	27	48	56	18	41	30	36	1077
·4714	37	33	27	31	61	35	44	30	47	49	26	26	15	36	50	39	22	58	50	55	30	55	27	29	34	44	48	46	1084
·8165	8	10	5	11	6	4	14	3	3	11	5	3	7	6	4	4	8	12	10	12	4	12	7	5	22	8	12	10	226
·9428	9	19	39	3	8	11	10	11	4	3	8	4	8	20	5	6	12	10	9	3	12	5	16	9	15	7	8	6	280
1·2472	—	1	—	—	1	1	2	—	1	2	—	—	2	—	2	—	4	—	—	—	—	1	1	1	8	—	2	2	31
1·4142	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	2	—	—	—	5
1·6330	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
Totals	100	100	100	68	100	100	100	100	100	100	100	100	60	100	100	76	100	100	100	100	100	100	100	100	100	100	100	100	2704

TABLE XXXVII—N 1908, A—C.

Locules per Ovary in three Series.

Tree	A				B				C			
	2	*3*	*4*	*5*	*2*	*3*	*4*	*5*	*2*	*3*	*4*	*5*
11	32	100	1	—	14	100	—	—	5	104	—	—
12	—	100	—	—	—	100	—	—	—	109	1	—
13	—	50	1	—	—	100	3	—	2	118	—	—
14	9	50	—	—	10	50	—	—	14	75	—	—
15	—	100	—	—	1	100	—	—	—	108	2	—
16	—	90	—	—	5	90	—	—	—	121	1	—
17	—	100	—	—	—	100	—	—	—	109	1	—
18	—	50	—	—	2	100	6	—	1	111	3	—
19	5	90	—	—	6	80	—	—	11	153	—	—
20	—	80	—	—	—	90	—	—	—	111	—	—
21	1	50	—	—	—	85	1	—	—	103	2	—
22	—	55	3	—	—	65	1	—	—	102	1	—
23	—	50	2	—	—	50	1	—	—	61	1	—
24	—	60	5	—	2	65	—	—	—	110	—	—
25	—	65	—	—	—	100	1	—	—	163	4	—
26	—	50	—	—	—	50	—	—	—	76	—	—
27	—	65	—	—	—	80	—	—	—	122	—	—
29	—	100	—	—	—	100	4	—	—	123	2	—
*30**	—	80	—	—	—	85	1	—	—	139	2	—
31	—	100	2	—	—	100	—	—	1	105	3	—
32	—	65	3	1	—	100	9	—	—	184	26	—
33	5	90	—	—	2	85	—	—	13	134	—	—
34	—	50	3	—	—	90	15	2	—	118	16	—
35	—	100	10	—	—	100	13	—	—	126	18	—
37	—	100	—	—	—	100	—	—	—	124	—	—
38	2	100	—	—	18	100	—	—	3	153	—	—
39	—	40	—	—	—	100	3	—	1	135	3	—
40	—	65	—	—	—	100	—	—	1	158	—	—
	54	2095	30	1	60	2465	58	2	52	3355	86	—

* Shrub 30, Series C has one fruit intermediate between 3- and 4-loculed condition.

TABLE XXXVIII—L 1906.

Ovules per Locule.

	5	*6*	*7*	*8*	*9*	*10*	*11*	Totals
6—10	12	127	142	214	23	4	—	522
11—15	2	104	393	810	76	24	1	1410
16—20	2	28	168	553	53	15	—	819

TABLE XXXIX—O 1906.

Total Ovules per Ovary.

	17	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	Totals
A ...	7	15	19	18	21	21	21	35	8	4	5	—	—	—	174
B ...	—	6	11	32	40	77	107	146	22	10	7	4	7	1	470
C ...	—	1	3	3	18	41	54	119	15	6	6	5	1	1	273

TABLE XL—A 1906.

	A		B		C	
	f	%	*f*	%	*f*	%
·0000	59	33·908	177	37·659	126	46·154
·4714	85	48·850	244	51·914	115	42·124
·8165	17	9·770	23	4·893	19	6·960
·9428	12	6·896	24	5·106	11	4·029
1·2472	1	·574	2	·425	2	·733
Totals	174	—	470	—	273	—

TABLE XLI—ALC 1908, A.

Coefficient of Asymmetry.

Locular Composition.	*·0000*	*·4714*	*·8165*	*·9428*	*1·2472*	*1·4142*	*1·6330*	*1·6997*	*1·8856*	Totals
3 even ...	462	—	—	130	—	—	2	—	1	595
2 even, 1 odd	—	614	138	—	21	14	—	1	—	788
1 even, 2 odd	—	443	95	—	22	8	—	5	—	573
3 odd ...	103	—	—	35	—	—	1	—	—	139
Totals ...	565	1057	233	165	43	22	3	6	1	2095

TABLE XLII—ALC 1908, B.

Coefficient of Asymmetry.

Locular Composition.	*·0000*	*·4714*	*·8165*	*·9428*	*1·2472*	*1·4142*	*1·6330*	*1·6997*	*1·8856*	*2·0548*	*2·1602*	Totals
3 even ...	610	—	—	130	—	—	—	—	1	—	—	741
2 even, 1 odd	—	739	146	—	35	9	—	6	—	1	—	936
1 even, 2 odd	—	451	142	—	28	7	—	2	—	2	1	633
3 odd ...	110	—	—	42	—	—	3	—	—	—	—	155
Totals ...	720	1190	288	172	63	16	3	8	1	3	1	2465

TABLE XLIII—ALC 1908, C.

Coefficient of Asymmetry.

Locular Composition.	*·0000*	*·4714*	*·8165*	*·9428*	*1·2472*	*1·4142*	*1·6330*	Totals
3 even ...	1001	—	—	263	—	—	1	1265
2 even, 1 odd	—	744	160	—	20	4	—	928
1 even, 2 odd	—	340	66	—	11	1	—	418
3 odd ...	76	—	—	17	—	—	—	93
Totals ...	1077	1084	226	280	31	5	1	2704

TABLE XLIV—AL 1908, B.

Ovules per Locule.

Coefficient of Asymmetry.	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	Totals
·0000	—	—	—	18	555	255	1242	57	33	—	—	2160
·4714	—	—	5	169	652	1089	1191	381	81	2	—	3570
·8165	—	—	7	48	161	248	232	126	33	8	1	864
·9428	—	—	7	26	137	66	206	30	40	4	—	516
1·2472	—	—	4	19	31	36	44	28	18	8	1	189
1·4142	—	—	1	6	3	6	16	6	5	5	—	48
1·6330	—	—	—	1	—	3	—	3	—	2	—	9
1·6997	1	1	—	1	4	3	4	4	2	1	3	24
1·8856	—	—	—	—	2	—	—	—	1	—	—	3
2·0548	—	—	—	2	1	2	1	—	2	1	—	9
2·1602	—	—	—	—	1	1	—	—	—	1	—	3
Totals	1	1	24	290	1547	1709	2936	635	215	32	5	7395

TABLE XLV—LCL 1908, B.

Ovules per Locule.

Locular Composition.	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	Totals
3 "even" ...	—	—	7	—	694	—	1448	—	74	—	—	2223
2 "even," 1 "odd"	1	1	15	104	610	607	1144	214	97	10	5	2808
1 "even," 2 "odd"	—	—	2	141	243	778	344	331	44	16	—	1899
3 "odd" ...	—	—	—	45	—	324	—	90	—	6	—	465
Totals	1	1	24	290	1547	1709	2936	635	215	32	5	7395

A BIOMETRIC STUDY OF PHAGOCYTOSIS WITH SPECIAL REFERENCE TO THE "OPSONIC INDEX."

SECOND MEMOIR. ON THE DISTRIBUTION OF THE MEANS OF SAMPLES.

By M. GREENWOOD, Junr., and J. D. C. WHITE.

(*From the Statistical Laboratories of the Lister Institute of Preventive Medicine and the London Hospital.*)

In a recent memoir* we communicated the results of an analysis undertaken with the object of throwing light on difficulties associated with the ordinary method of estimating opsonic indices. The main results of that investigation can be summarised in the following way:

(1) Phagocytic counts have pronouncedly skew frequency distributions and are good examples of Pearson's skew curves.

(2) The means of small samples have likewise a skew distribution.

(3) The use of a thicker bacillary emulsion while diminishing the skewness, does not, under possible experimental conditions, do so sufficiently to allow of testing on the basis of a Gaussian curve.

With respect to (2), we pointed out that results based, as were ours, upon samples extending in no case to more than 2000 cells, were provisional and that we intended to re-investigate this part of the subject with the help of more adequate material.

The present memoir comprises a study of the distribution of a large sample of phagocytic cells, the distributions of the means of sub-samples and some preliminary contributions to the problem of mean distributions in general.

The basis of this work is a count of 20,000 cells made by one of us (J. D. C. W.) who had had 18 months' previous experience in the Inoculation Department of the London Hospital.

* *Biometrika*, Vol. vi. p. 377.

In undertaking this large count, those methods were adopted which are usually advised in the preparation of films. The mixture, to be subsequently placed on the slides, was made up of three equal parts, (1) serum, (2) corpuscles, (3) bacterial emulsion. The first two were derived from the operator and, in the case of the corpuscles, care was taken that they should be equally distributed, the upper portion of the centrifugalised blood being taken off and thoroughly mixed in a small tube. The bacterial emulsion was prepared as usual and carefully mixed so as to secure the greatest possible homogeneity. Since we desired a large count of a homogeneous character, not a comparison of different mixtures, it was decided to use one pipette only, but of larger calibre than usual, in order to secure the greater amount of mixture requisite for the preparation of a large number of slides. This pipette was throttled for ease of manipulation. The proportions of the three materials were carefully mixed and the pipette placed in an incubator for fifteen minutes, the emulsion having been designed to give an average of three to four bacilli per cell. The pipette was slightly rotated from time to time in order to keep the corpuscles from settling. Sufficient slides being in readiness, the pipette was withdrawn from the incubator, the contents again thoroughly mixed, and films were prepared as quickly as possible. These were then stained with Aniline Fuchsin and Methylene Blue; thereafter cover-slips were applied with Canada Balsam to preserve the slides during the process of counting. The general quality both of films and staining was exceedingly good, very few slides being of an inferior character.

The counting of so large a number of cells as 20,000 was, needless to say, sufficiently laborious, especially in regard to the leucocytes containing more than 10 bacilli apiece. Reasonable accuracy was found possible up to 15 and the few cells containing more than this were marked 16. A mechanical stage was used—the only method, in the operator's opinion, which can secure that the same cell shall not be counted twice over. The only cells omitted were those the outline of which was indistinct, or where for any reason, e.g. the clumping of cells or bacilli, it could not be determined how many bacilli lay in an individual cell; the possible error introduced by this latter criterion will be discussed below.

Definite fragments of bacteria were counted as bacteria; so also were bacilli in definite contact with the periphery of cells. In order to avoid mental prejudice, addition of the rows of figures was postponed generally for days and always at least until the end of the day's work. In this way 20,000 cells were counted in the course of about a month, with occasional intervals. The sheets of figures were then taken and the totals of successive twenty-fives, fifties and hundreds were determined; then the frequency of each number per cell on each sheet of 500 cells was tabulated, this latter process incidentally eliminating any arithmetical mistakes. The frequencies on each of the forty sheets were then added together and the total frequency obtained.

The actual distribution of the twenty thousand cells is set forth in Table I. Proceeding to fit a curve, without using Sheppard's corrections, the constants of

Table II were obtained and Graph 1 shows the theoretical curve and observations. It is evident that, although the curve represents the general trend of the observations with fair accuracy, the goodness of fit as evaluated in the ordinary way is poor. This matter appears to require comment and explanation.

TABLE I.

Actual Frequencies of 20,000 *Cells, together with Corrected Values determined on the Basis of a Random Count of* 2,000 *Cells from the same Material.* (See p. 509 *infra.*)

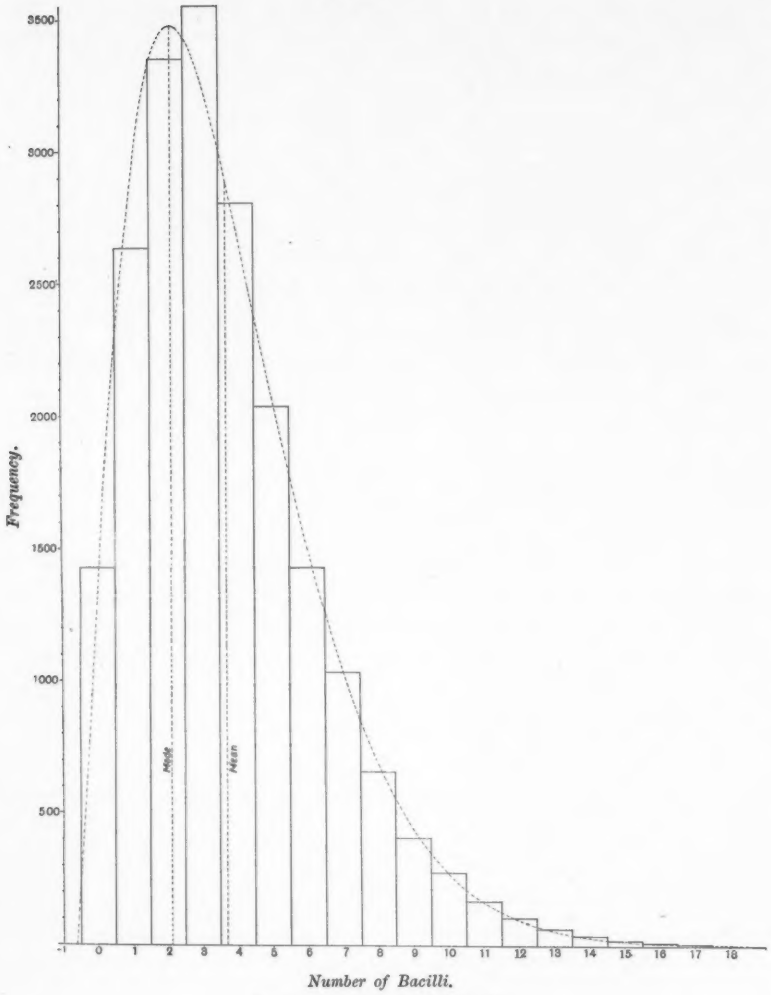
Number of Bacilli per Cell	Original Value	Corrected Value*
0	1428	1402·75
1	2633	2636·57
2	3351	3355·55
3	3556	3560·83
4	2807	2810·81
5	2038	2040·77
6	1430	1431·94
7	1031	1032·40
8	658	658·89
9	407	407·55
10	271	271·37
11	163	163·22
12	103	103·14
13	61	61·08
14	37	37·05
15	17	17·02
16	9	9·01
Totals	20000	19999·95

It will have been noticed that for the purposes of this count we did not enumerate cells which contained clumps of bacteria. Our reason for this omission was that, to the best of our knowledge, workers ordinarily disregard cells the contents of which cannot be resolved into definite bacilli or fragments of bacilli, and we were most anxious to use data so far as possible identical with those employed in actual practice. It was, however, subsequently pointed out to us, that this process was open to serious objection for the following reasons. In the first place, some workers of great reputation never employ for diagnostic purposes mixtures containing an appreciable number of clumps, say more than two or three per cent. In the second place, the exclusion of clumps would tend arbitrarily to heighten the proportion of cells containing 0 bacilli, since cells containing clumps must be regarded as possessing a phagocytic power at least equal to those which contain one bacillus apiece; indeed, Harvey and McKendrick have adduced evidence† which supports a belief that it is rather the number of bacteria than the number of acts of ingestion which should be taken as a measure of phagocytic power.

* 1428 cells with 0 bacilli per cell will occur not in 20,000, but in 20,360; hence in 20,000 we shall have not 1428 but 1402·75, the excess is distributed proportionally among the other groups: see p. 509.

† *Biometrika*, Vol. VII. p. 64.

TABLE II.



GRAPH 1. Curve of 20,000 cells.

Constants of Fitted Curve.

$\sigma=2{\cdot}6031$, $\beta_1=1{\cdot}0365$, $\beta_2=4{\cdot}2622$, $\kappa_2=-1{\cdot}6752$, $r=22{\cdot}8160$, $b=41{\cdot}5558$, $m_1=1{\cdot}3806$, $m_2=19{\cdot}4354$, $a_1=2{\cdot}7561$, $a_2=38{\cdot}7997$, $y_0=3478{\cdot}81$, Skewness $=\cdot6069$, Mode at $2{\cdot}1000$, Mean at $3{\cdot}6797$, Start of curve at $-\cdot6562$, End at $40{\cdot}8996$, $\chi^2=86{\cdot}2$, P very small,

$$y=3478{\cdot}81\left(1+\frac{x}{2{\cdot}7561}\right)^{1{\cdot}3806}\left(1-\frac{x}{38{\cdot}7997}\right)^{19{\cdot}4354}.$$

Recognising the force of this objection, it was decided to examine the effect of including clumps in the count. To this end, another 2000 cells were counted on the slides already used and clumped cells were included. Of the 2000 cells counted in this way 36, or 1·8 per cent., contained clumped bacteria. Now if we allow, as seems reasonable, that this latter enumeration was a fair sample of the population, then the frequencies in our 20,000 cells would be changed in the manner shown in the second column of Table I. We have distributed the clumped cells among the other frequencies above zero in accordance with the proportions of those frequencies in the original count, since the small number of clumped cells which were present in the 2000 did not allow us to determine the true proportion of clumps, whether mainly of two, three, or more bacilli*.

The alteration which is effected by this correction is so trifling, that our erroneous method of counting, if it be considered an erroneous method, can hardly have been the cause of the poor fit which resulted, and we have not thought ourselves justified in discarding the data originally collected. Some other source of the poor fit must be looked for.

Other possible explanations are:

(1) The existence of heterogeneity in the material dependent on the fact that cells from different parts of the films are not, or may not be, strictly comparable with each other.

(2) An artificial heterogeneity dependent on the process of counting, other than that already discussed.

It has been noticed before that a large sample of material, even when adequately described by a frequency curve from the diagrammatic point of view, fails to satisfy the approved test. Indeed, many statisticians have been, as Pearson has remarked, far too easily satisfied with the test of mere inspection.

Elderton writes: "I have found in applying the test, that when numbers dealt with are very large, the probability is often small, even though the curve appears to fit the statistics very closely. The explanation is that the statistics with which we deal in practice nearly always contain a certain amount of extraneous matter, and heterogeneity is concealed in a small experience by the roughness of the data. The increase in the number of cases observed removes the roughness, but the heterogeneity remains. The meaning, from the curve-fitting point of view, is that the experience is really made up of more than one frequency curve, but a certain curve, approximating to the one calculated, predominates†."

It will have been noticed that the poverty of fit is mainly due to the cells containing one bacillus being in defect and those containing three in excess; these two groups have added nearly 58 to the value of χ^2. Now the work of counting is excessively monotonous, and after going through some hundreds, it seems impossible to escape an impression that a certain measurement, say 3 bacilli per cell, is modal. Hence a tendency will arise to place any doubtfuls within that particular class. If the count be limited to one or two thousands, the error so introduced may not appreciably affect the result, but it will do so if the data run to many thousands, since the same percentage deviation in a large as in a small experience has naturally a much greater influence on the fit.

* The 36 cells which contained "clumps" were actually distributed as follows:

Number of Bacilli	Number of Cells	Number of Bacilli	Number of Cells	Number of Bacilli	Number of Cells
1	2	4	4	7	5
2	6	5	8	8	1
3	8	6	0	9	2

† *Frequency Curves and Correlation*, by W. Palin Elderton, p. 142.

The only prejudice to which the operator seems to have been—and thinks he was—subject was the idea which impressed itself on his mind during the count that the number of 3 bacilli per cell occurred with the greatest frequency; this idea may have inclined him to welcome threes with undue alacrity. It is possible that this may have helped in some degree towards producing an excess of threes in the final result. The operator believes that, at least in his case, the auditory sense is more impressed with the sound of numbers containing *e*'s and *i*'s, 3, 5, 9, than with others; he thus tended to believe that these numbers occurred more frequently than was actually the case. Evidently the truth of the matter cannot be exactly known until several independent observers have undertaken the same large count.

We have already mentioned the difficulty regarding clumps; another point is as to whether some gross heterogeneity were produced either by notable differences in various slides or by variations in the method of counting on different occasions, differences of which the operator was unconscious. In making the frequency tables just considered, there emerged necessarily a rough representation of a frequency curve for each five hundred cells. All these curves showed a rough similarity and each one a rough regularity. In some cases, however, the regularity was appreciably less than in others. On going through the data a second time, the groups of 500 which showed the greatest irregularities were excluded, on the supposition that they might represent a somewhat lower grade of accuracy in counting and might have introduced a gross heterogeneity into the total count. The application of this criterion led to the exclusion of 5000 cells. The remaining 15,000 cells were then analysed as before.

Tables III, III A and Graph 2 indicate the results of this process. There is a slight improvement in the fit, but the change is not pronounced. It can hardly be said that the results justify the exclusion.

TABLE III.

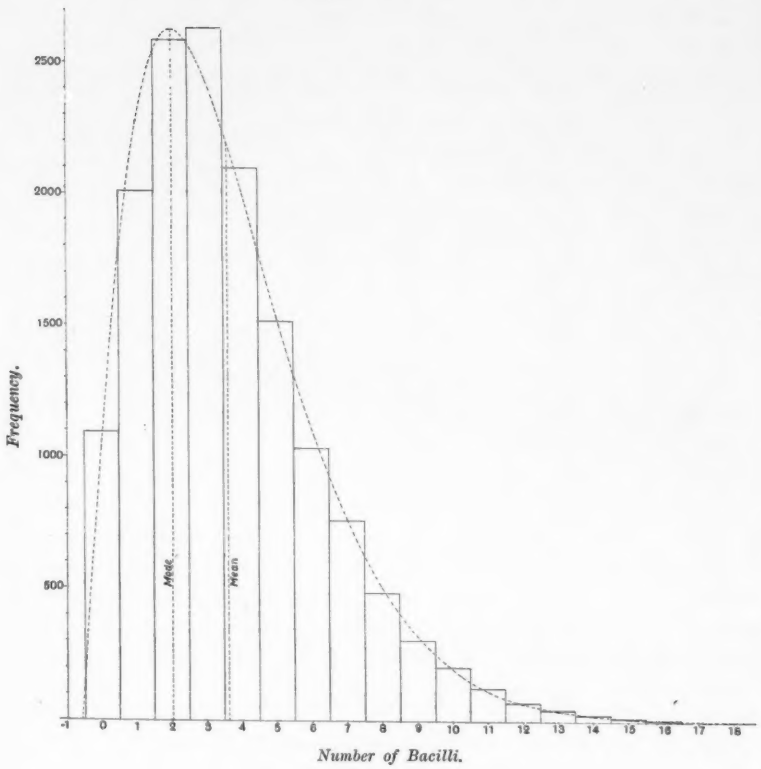
Actual Distribution of 15,000 *Cells.*

Bacilli per Cell	Number of Cells	Bacilli per Cell	Number of Cells
0	1092	*9*	306
1	2008	*10*	203
2	2585	*11*	126
3	2632	*12*	69
4	2096	*13*	44
5	1518	*14*	24
6	1035	*15*	12
7	760	*16*	6
8	484		

Although we should naturally have preferred to obtain closer fits, the result does not seriously affect the value of our material for the purposes of the present investigation. It is clear that the heterogeneity is not caused by a compounding of Gaussian distributions, and that we have to deal with markedly skew variation. We sincerely trust that some other workers may be tempted to repeat the experiment. We now turn to our main inquiry, viz. the problem of sampling.

The 20,000 cells were arranged in *consecutive* 25's, 50's and 100's, and grouped as shown in Table IV. In each case the group is an inclusive one, thus 76—83

TABLE III A



GRAPH 2. Curve of 15,000 cells.

Constants of Fitted Curve.

β_1	β_2	κ_2	r	b	m_1	m_2	a_1	a_2
1·0381	4·2346	−1·5233	20·4319	38·1013	1·2784	17·1535	2·6426	35·4587

y_0	Skewness	Mode	Mean	Start	End	P
2626·615	·61999	2·0360	3·6421	−·5533	37·4947	·00001

$$y = 2626{\cdot}615\left(1+\frac{x}{2{\cdot}6426}\right)^{1{\cdot}2784}\left(1-\frac{x}{35{\cdot}4587}\right)^{17{\cdot}1535}.$$

TABLE IV.

Actually Observed Frequencies of the Samples from 20,000 *Cells.*

Samples of 25		Samples of 50		Samples of 100	
Number of Bacilli per Sample	Number of Samples	Number of Bacilli per Sample	Number of Samples	Number of Bacilli per Sample	Number of Samples
44— 51	1	*112—123*	2	*267—282*	2
52— 59	14	*124—135*	9	*283—298*	9
60— 67	38	*136—147*	21	*299—314*	4
68— 75	74	*148—159*	30	*315—330*	13
76— 83	128	*160—171*	58	*331—346*	29
84— 91	145	*172—183*	83	*347—362*	36
92— 99	147	*184—195*	71	*363—378*	34
100—107	113	*196—207*	64	*379—394*	34
108—115	74	*208—219*	31	*395—410*	15
116—123	37	*220—231*	15	*411—426*	11
124—131	19	*232—243*	10	*427—442*	5
132—139	7	*244—255*	3	*443—458*	3
140—147	2	*256—267*	1	*459—474*	1
148—155	0	*268—279*	0	*475—490*	1
156—163	0	*280—291*	1	*491—506*	2
164—171	0	*292—303*	1	*507—522*	1
172—179	1				
Totals	800	—	400	—	200

includes samples with 76 and samples with 83 bacilli per 25 cells. For convenience of calculation, we took an arbitrary origin and unit of grouping in each case. In obtaining the moment coefficients Sheppard's corrections were used in all cases; Table V gives the constants and equations, Graphs 3, 4, 5 illustrate the curves.

TABLE V.

Curves Resulting from an Analysis of the 20,000 *Cells.*

Material	Number of Samples	Mean (in terms of Bacilli per Sample)	Mode (in terms of Bacilli per Sample)	σ (in terms of Bacilli)	β_1	β_2
Samples of 25	800	92·04	89·7488	16·7184	·1033	3·4321
Samples of 50	400	184·07	181·3964	25·7496	·1467	4·0907
Samples of 100	200	367·62	360·2632	39·9824	·3004	4·3500

Material	κ_2	r	m	ν	a	β_2-3	Skewness	P (Fit)
Samples of 25	·1436	25·2068	13·6033	−10·3220	9·5149	·4321	·1371	·9765
Samples of 50	·0665	10·1436	6·0718	− 2·7070	6·2691	1·0907	·1066	·8673
Samples of 100	·1367	10·1723	6·0862	− 4·0483	7·0316	1·3500	·1840	·1727

Equations: $$y = 23{\cdot}428\left[1+\frac{x^2}{9{\cdot}5149^2}\right]^{-13{\cdot}6034} \times e^{10{\cdot}3220\tan^{-1}\frac{x}{9{\cdot}5149}}.$$

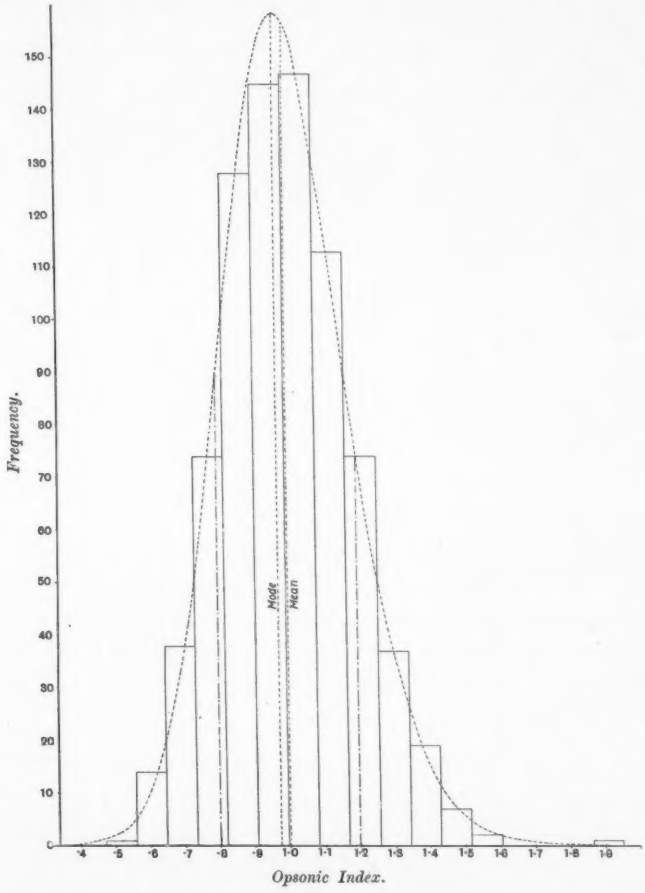
Origin at 2·6712, where 39·5 bacillj are taken as 0 and 8 bacilli per 25 cells form the statistical unit of grouping.

$$y = 60{\cdot}047\left[1+\frac{x^2}{6{\cdot}2691^2}\right]^{-6{\cdot}0718} \times e^{2{\cdot}7070\tan^{-1}\frac{x}{6{\cdot}2691}}.$$

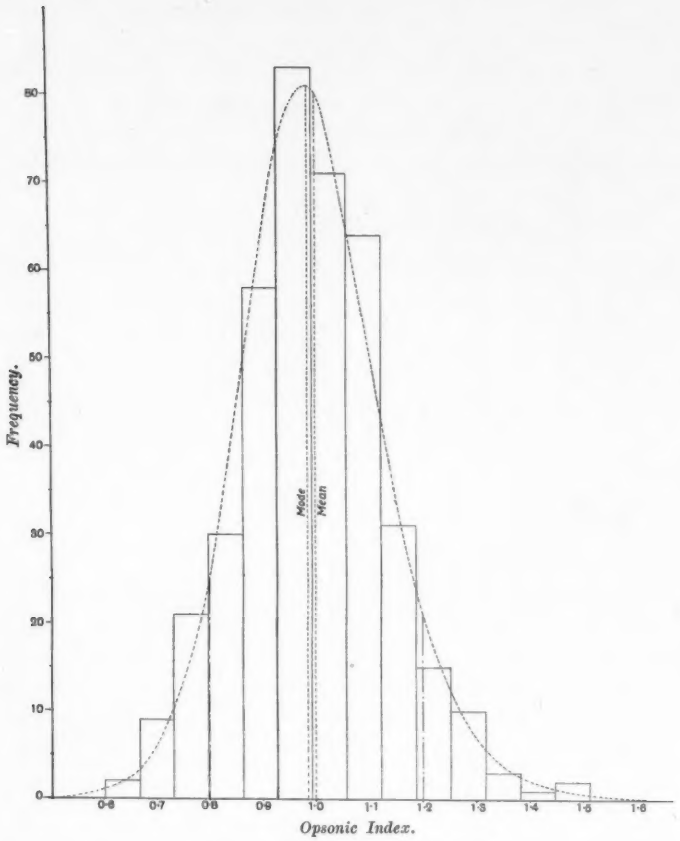
Origin at 4·87446, 105·5 bacilli taken as 0 and 12 bacilli per 50 cells as the statistical unit of grouping.

$$y = 18{\cdot}117\left[1+\frac{x^2}{7{\cdot}0316^2}\right]^{-7{\cdot}0316} \times e^{4{\cdot}0483\tan^{-1}\frac{x}{7{\cdot}0316}}.$$

Origin at 4·021598, 258·5 bacilli taken as 0 and 16 bacilli per 100 cells as the statistical unit of grouping.



GRAPH 3. Curve of Samples of 25.



GRAPH 4. Curve of 400 Samples of 50.

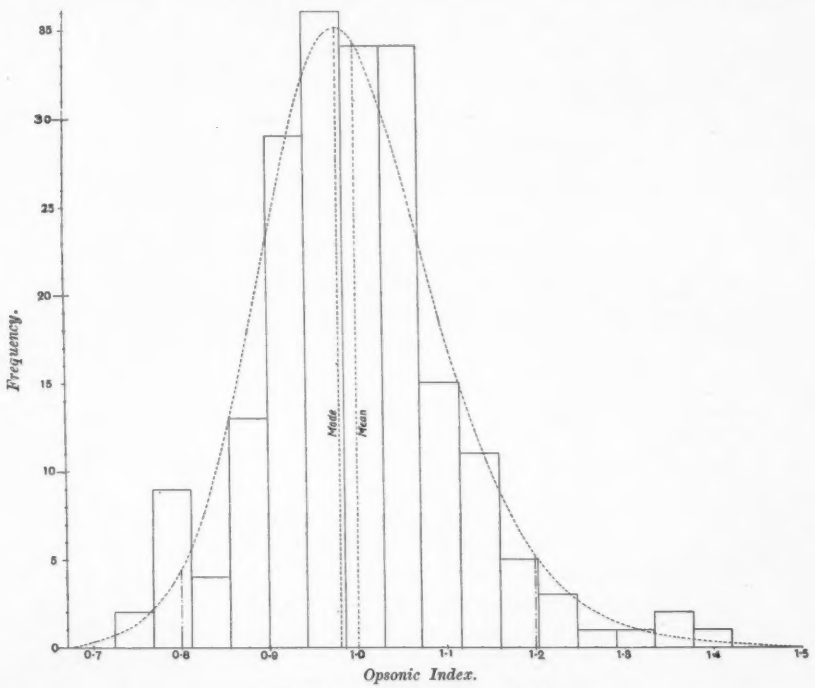
The figures below the abscissa represent opsonic indices in terms of the mean of the whole series. That is to say, each group unit was translated back into physical terms and then divided by the mean number of bacilli per sample for the whole count. For instance, in Graph 3, the abscissa ·9 was really 5 in statistical units corresponding to 39·5 (the origin) + 5 × 8 bacilli per 25 cells. The mean is 92·04, therefore the opsonic index in terms of the mean corresponding to this abscissa is $\frac{79{\cdot}5}{92{\cdot}04} = {\cdot}86$ *.

* The indices as plotted are only given to the nearest first decimal.

Objections may fairly be urged against this process. As we have previously remarked, the mean is theoretically not the best constant to use for descriptive purposes in the case of markedly skew frequencies and we suggested the mode as being more satisfactory. Harvey and McKendrick, in an important contribution to the subject*, have alluded to this point and remark with justice that the determination of the true mode requires calculation beyond the range of a laboratory worker. We have considered whether an approximation of sufficient accuracy might not be found. Pearson has shown† that in many cases a good enough value for the mode is given by:

Mode − Median = Twice the Distance from Mean to Median.

Unfortunately, however, the median cannot be determined with sufficient accuracy in the case of discontinuous variates such as phagocytic counts. Thus, we have experimentally no finer measurement than the integral number of bacilli per leucocyte; if we mass in a group all the m individuals containing the same number of bacilli and know that the median is the nth individual in that group, all we can do is to add $\frac{n}{m}$ of the unit of grouping to the next lower integral group measure. This is not nearly close enough in the case of such skew distributions; and we found that the value of the mode thus obtained differed materially from the real mode. It was not therefore possible to use this process with success in the case of our sample curves.



GRAPH 5. Curve of Samples of 100.

* *Biometrika*, Vol. VII. p. 64.

† Pearson, *ibid.* Vol. I. p. 260.

We next treated, for purposes of comparison, the selected count of 15,000 cells in a similar fashion. Table VI contains the actual frequencies, Table VI A the constants and equations. Let us now consider in detail points suggested by these analytical results. The values of P show that in all cases the fit is close except in that of the sample hundreds. It appeared to us probable that in this case we had adopted too fine a unit of grouping, so that we re-worked, using a coarser unit which gave 11 groups, with the following result,

$$y = 29{\cdot}9734\left(1 + \frac{x^2}{(4{\cdot}4732)^2}\right) \times e^{3{\cdot}7028 \tan^{-1}(x/4{\cdot}4732)},$$

which gives $P = {\cdot}232$.

This finding confirms the suggestion to some extent, but it must be admitted that the 100's are still definitely less satisfactory than the smaller samples; the explanation of this is not clear. Table VI B, which gives all the actual and calculated frequencies reduced to a common scale, will show the reader how little change was effected by the rejection of 5000 cells.

TABLE VI.

Actually Observed Frequencies of the Samples from 15,000 *Cells.*

Samples of 25		Samples of 50		Samples of 100	
Number of Bacilli per Sample	Number of Samples	Number of Bacilli per Sample	Number of Samples	Number of Bacilli per Sample	Number of Samples
44— 51	1	*112—123*	1	*267—282*	2
52— 59	11	*124—135*	8	*283—298*	8
60— 67	29	*136—147*	15	*299—314*	3
68— 75	62	*148—159*	25	*315—330*	10
76— 83	98	*160—171*	48	*331—346*	21
84— 91	115	*172—183*	66	*347—362*	27
92— 99	107	*184—195*	50	*363—378*	30
100—107	78	*196—207*	47	*379—394*	27
108—115	56	*208—219*	22	*395—410*	10
116—123	23	*220—231*	9	*411—426*	6
124—131	15	*232—243*	6	*427—442*	3
132—139	4	*244—255*	1	*443—458*	1
140—147	1	*256—267*	1	*459—474*	0
		268—279	0	*475—490*	0
		280—291	1	*491—506*	2
Totals	600	—	300	—	150

TABLE VI A.

Curves Resulting from an Analysis of 15,000 Selected Cells.

Material	Number of Samples	Mean (in terms of Bacilli per Sample)	Mode (in terms of Bacilli per Sample)	σ (in terms of Bacilli)	β_1	β_2
Samples of 25	600	91·1133	89·0117	16·3123	·0457	2·8211
Samples of 50	300	182·2200	179·4383	24·5984	·0980	3·7994
Samples of 100	150	363·5667	360·1649	37·0043	·1019	4·5418

Material	κ_2	r	m	ν	a	β_2-3	Skewness	P(Fit)
Samples of 25	−·0702	—	—	—	—	− ·1789	·1288	·9966
Samples of 50	+·0582	12·4217	7·2109	−3·0885	6·7230	+ ·7994	·1131	·7365
Samples of 100	+·0292	7·4301	2·7151	−1·2890	5·7783	+1·5418	·0920	·2164

Equations: $$y=114{\cdot}724\left(1+\frac{x}{7{\cdot}1842}\right)^{7{\cdot}0048}\left(1-\frac{x}{12{\cdot}8362}\right)^{12{\cdot}5157}.$$

Unit of grouping 8 bacilli per 25 cells. 39·5 bacilli per 25 cells taken as zero.

$$y=45{\cdot}045\left[1+\frac{x^2}{6{\cdot}7230^2}\right]^{-7{\cdot}2109}\times e^{3{\cdot}0885\tan^{-1}\frac{x}{6{\cdot}7230}}.$$

Origin at 4·7217, where 12 bacilli per 50 cells form the unit of grouping and the zero is at 105·5.

$$y=26{\cdot}445\left[1+\frac{x^2}{5{\cdot}7783^2}\right]^{-4{\cdot}7151}\times e^{1{\cdot}2890\tan^{-1}\frac{x}{5{\cdot}7783}}.$$

Origin at 5·5642, where 16 bacilli per 100 cells are the unit and the zero is at 258·5.

Next as to the agreement between the theoretical constants of the curves of means and their empirical values as above determined. Table VII gives the calculated values of B_1 and B_2-3 [assuming that where n is the number of individuals in the sample and β_1, β_2 the constants of the original distribution $B_1=\frac{\beta_1}{n}$ and $B_2-3=\frac{\beta_2-3}{n}$] together with the actual values.

In only one case, that of the 25's from the selected 15,000, is the agreement at all possible. "Student," in a very interesting communication*, has suggested that the divergence noted by him in our previous curve of 80 means, depended on the existence of homotyposis in the samples. This suggestion no doubt contains some truth and could be satisfactorily tested by mixing up our individual records of leucocytes and drawing really random samples. We have not felt justified in delaying publication of our results until we could find leisure to undertake this very laborious task, particularly as other factors are involved. This is clear from the fact that the divergence is more marked in the case of the larger samples which must be less influenced by homotyposis.

* *Biometrika*, Vol. VII. p. 210.

TABLE VI B.

Comparison of the Observed and Calculated Frequencies for Samples of 20,000 *and* 15,000 *Cells. All the Frequencies are reduced to* 100.

Twenty-Fives					Fifties					Hundreds				
Bacilli	20,000 Count		15,000 Count		Bacilli	20,000 Count		15,000 Count		Bacilli	20,000 Count		15,000 Count	
	Observed	Calculated	Observed	Calculated		Observed	Calculated	Observed	Calculated		Observed	Calculated	Observed	Calculated
44— 51	·13	·35	·17	·22	*112—123*	·5	·5	·33	·49	*267—282*	1	1·0	1·33	1·09
52— 59	1·75	1·40	1·83	1·51	*124—135*	2·25	1·54	2·67	1·61	*283—298*	4·5	1·86	5·33	1·95
60— 67	4·75	4·22	4·83	4·75	*136—147*	5·25	4·17	5·0	4·49	*299—314*	2	4·52	2·0	4·49
68— 75	9·25	9·47	10·33	10·74	*148—159*	7·5	9·24	8·33	9·99	*315—330*	6·5	8·97	6·67	8·91
76— 83	16·0	15·71	16·33	16·20	*160—171*	14·5	15·71	16·0	16·73	*331—346*	14·5	14·03	14·0	14·79
84— 91	18·12	19·44	19·17	18·89	*172—183*	20·75	19·80	22·0	20·52	*347—362*	18	17·14	18·0	18·72
92— 99	18·38	18·44	17·83	17·75	*184—195*	17·75	18·63	16·67	18·65	*363—378*	17	16·68	20·0	18·03
100—107	14·13	13·89	13·0	13·73	*196—207*	16·0	13·63	15·67	13·10	*379—394*	17	13·40	18·0	13·61
108—115	9·25	8·63	9·33	8·80	*208—219*	7·75	8·23	7·33	7·52	*395—410*	7·5	9·28	6·67	8·50
116—123	4·63	4·61	3·83	4·65	*220—231*	3·75	4·36	3·0	3·75	*411—426*	5·5	5·77	4·0	4·68
124—131	2·38	2·18	2·50	2·00	*232—243*	2·5	2·13	2·0	1·71	*427—442*	2·5	3·34	2·0	2·39
132—139	·88	·95	·67	·68	*244—255*	·75	1·00	·33	·74	*443—458*	1·5	1·85	·66	1·18
140—147	·25	·39	·17	·21	*256—267*	·25	·46	·33	·50	*459—474*	·5	1·00	—	1·13
148—155	0	·15			*268—279*	·5	·39	0		*475—490*	·5	·54	—	
156—163	0	·09			*280—291*			·33		*491—506*	·1	·59	1·33	
164—171	0				*292—303*					*507—522*	·5			
172—179	·13													
Totals	100·03	99·92	99·99	100·13	—	100·00	99·79	99·99	99·80	—	100·00	99·97	99·99	99·47

TABLE VII.

Number in Sample	B_1 (empirical)	B_1 (calculated)	B_2-3 (empirical)	B_2-3 (calculated)
(20,000 Count)				
25	·1033	·0415	·4321	·0505
50	·1467	·0208	1·0907	·0252
100 (1st curve) ...	·3005	·0104	1·3501	·0126
100 (2nd curve) ...	·3142	·0104	1·4601	·0126
(15,000 Count)				
25	·0457	·0415	−·1789	·0494
50	·0980	·0208	·7994	·0247
100	·1019	·0104	1·5418	·0124

We can now examine the accuracy of the opsonic method in the way attempted in our former paper, viz. we can inquire what the chances are of obtaining from a "population" of means, samples giving different indices in terms of the real mean of all such samples. In similar cases, it is usual to divide the frequency curve into a series of equal areas, setting up ordinates on either side of the mean or mode. Since this elaborate graphical method requires the services of a specially skilled draughtsman, and considerably increases the cost of production, we decided to adopt the plan followed in our last paper. A considerable number of additional ordinates were calculated for each sample curve and we determined the area from the beginning or end of the curve up to specified ordinates which corresponded to indices of ·6, ·7, ·8, etc. in terms of the mean. From these areas the Tables of Chances (VIII, IX, X) were deduced.

TABLE VIII.

Chances of Obtaining Certain Deviations in Index Values.

Opsonic Index in terms of the Mean	Samples of 25	
	Fraction of the Total Area bounded by the corresponding ordinate (Total Area = 1)	Odds against the occurrence of such a Deviation or a Greater
·6	·0089	111 to 1
·7	·0396	24·3 to 1
·8	·1276	6·8 to 1
·9	·2973	2·4 to 1
1·1	·2740	2·7 to 1
1·2	·1324	6·6 to 1
1·3	·0558	16·9 to 1
1·4	·0211	46·4 to 1
Beyond the limits *1·4—·6*	·0300	32·3 to 1
" " *1·3—·7*	·0954	9·5 to 1
" " *1·2—·8*	·2600	2·9 to 1
" " *1·1—·9*	·5713	·75 to 1

TABLE IX.

Chances of Obtaining Certain Deviations in Index Values.

Opsonic Index in terms of the Mean	Samples of 50	
	Fraction of the Total Area bounded by the corresponding ordinate (Total Area = 1)	Odds against the occurrence of such a Deviation or a Greater
·7	·0112	88·3 to 1
·8	·0618	15·2 to 1
·9	·2283	3·4 to 1
1·1	·2160	3·6 to 1
1·2	·0770	12 to 1
1·3	·0248	39·3 to 1
Beyond the limits 1·3—·7	·0360	26·8 to 1
„ „ 1·2—·8	·1388	6·2 to 1
„ „ 1·1—·9	·4443	1·3 to 1

TABLE X.

Chances of Obtaining Certain Deviations in Index Values.

Opsonic Index in terms of the Mean	Samples of 100	
	Fraction of the Total Area bounded by the corresponding ordinate (Total Area = 1)	Odds against the occurrence of such a Deviation or a Greater
·7	·0006	1666 to 1
·8	·0216	45·4 to 1
·9	·1661	5·0 to 1
1·1	·1618	5·2 to 1
1·2	·0420	22·8 to 1
1·3	·0109	90·8 to 1
Beyond the limits 1·3—·7	·0115	86·6 to 1
„ „ 1·2—·8	·0636	14·7 to 1
„ „ 1·1—·9	·3279	2·1 to 1

If we adopt about a ten to one chance as the limiting value for evidence of differentiation, then the limits, for single determinations of the index, are in the case of 25's, roughly 1·3—·7; for 50's, 1·25—·75; for 100's, 1·2—·85. Everyone has his own standard of accuracy or reliability; all we mean is that unless a given

index is beyond these limits, the chances are less than ten to one against its being drawn from a normal population. We now come to a most important qualification of the value of these tables.

In the table as given, the measure is of the probable deviation of a sample from the real mean of the samples as determined from a long count. If the control mean, i.e. the denominator of the fraction which we call the opsonic index, has been determined from a large number of cells, say 1000, and the tested serum is measured on the basis of a small sample, then our results are directly applicable; but this supposition does not represent the usual state of affairs. In ordinary practice, two small samples, each of not more than 100 cells, serve as control and test, there is therefore, so to speak, a double source of variation; odds based on the results obtained above would over-estimate the reliability of any given determination. This is a point of considerable importance which has, we think, escaped the attention of writers on the subject; its elucidation would involve a careful examination of the whole theory of index distributions, and would

TABLE XI.

Constants of the Experimental Index Frequencies.

[Mean, Mode and σ are expressed in terms of the Index, i.e. in physical units.]

Material	Number of Samples	Mean	Mode	σ	β_1	β_2	κ_2	r
25's	800	1·0362	·9432	·2585	·7099	4·3914	·9619	—
50's*	400	1·0205	·9536	·2204	·6952	4·8945	·3635	11·2690
100's	200	1·0120	·9689	·1553	·3654	3·7138	·9032	—

Material	m	ν	a	β_2-3	Skewness	y	p	P (Fit)
25's	—	—	—	1·3914	·3600	326·1439	27·4967	·0853
50's	6·6345	−8·5166	5·6343	1·8945	·2912	—	—	·8953
100's	—	—	—	·7138	·2780	491·3142	48·7621	·9355

Equations: $y = \text{antilog } 43{\cdot}6116 \left(x^{-27{\cdot}4967} \times e^{-\frac{326{\cdot}1439}{x}} \right)$.

A change of ·1 in the index is taken as the statistical unit of grouping and index ·25 as zero.

$$y = 6{\cdot}2092 \left(1 + \frac{x^2}{5{\cdot}6343^2} \right)^{-6{\cdot}6345} \times e^{8{\cdot}5166 \tan^{-1} \frac{x}{5{\cdot}6343}}.$$

A change of ·1 in the index is taken as the statistical unit of grouping and index ·45 as zero. The origin is at 1·4468.

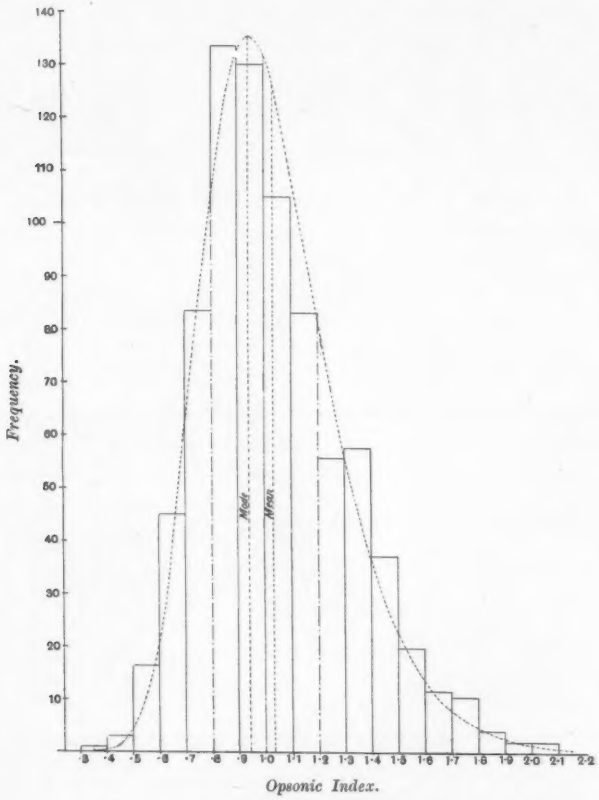
$$y = \text{antilog } 71{\cdot}8335 \left(x^{-48{\cdot}7621} \times e^{-\frac{491{\cdot}3142}{x}} \right).$$

A change of ·1 in the index is taken as the statistical unit of grouping and index ·45 as zero.

* A slight slip has been found in the values of β_1 and β_2 for this series, they should be ·6836 and 4·9855. This error does not, however, visibly alter the form of the curve in Graph 7.

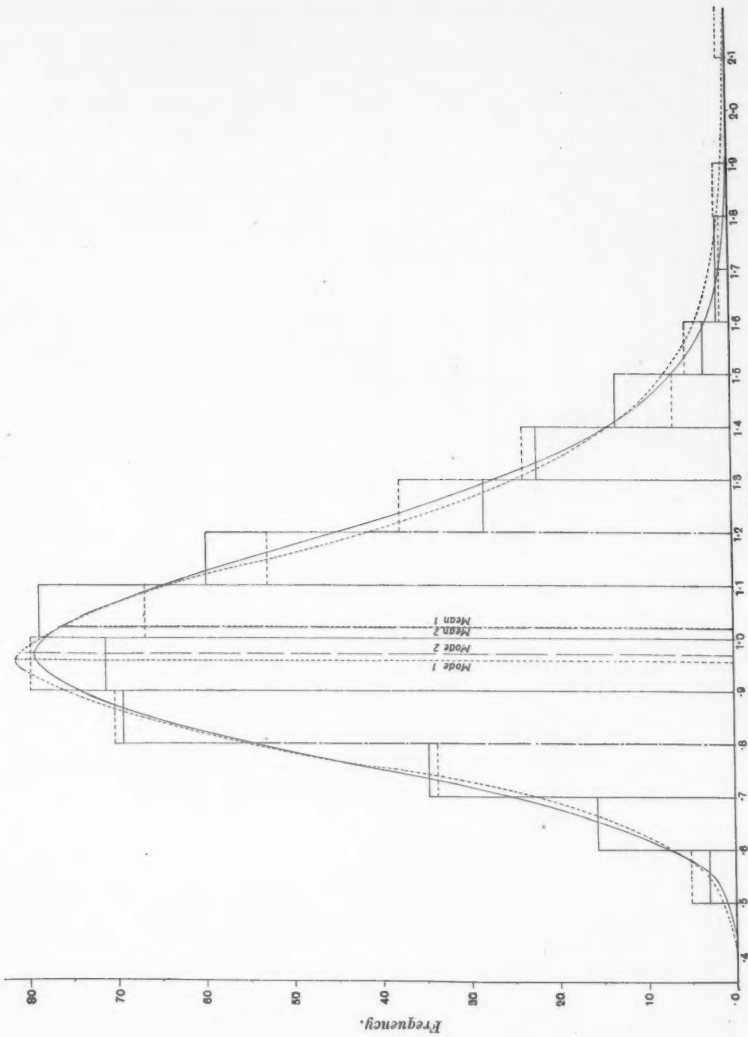
accordingly present features of interest quite apart from the particular case of the opsonic index. These remarks apply with special force to medical statistics which so frequently involve the use of ratios or indices.

The importance of the problem tempted us to devote some little time to a theoretical consideration of the subject and we arrived at a few analytical results which might, we thought, be of some value. It is, however, unnecessary for us to publish these incomplete investigations, since Professor Karl Pearson has succeeded in obtaining a full solution of the problem*. We therefore confine ourselves to an empirical solution which is sufficient for our immediate object.



GRAPH 6. Curve of 25's secondary without replacement.

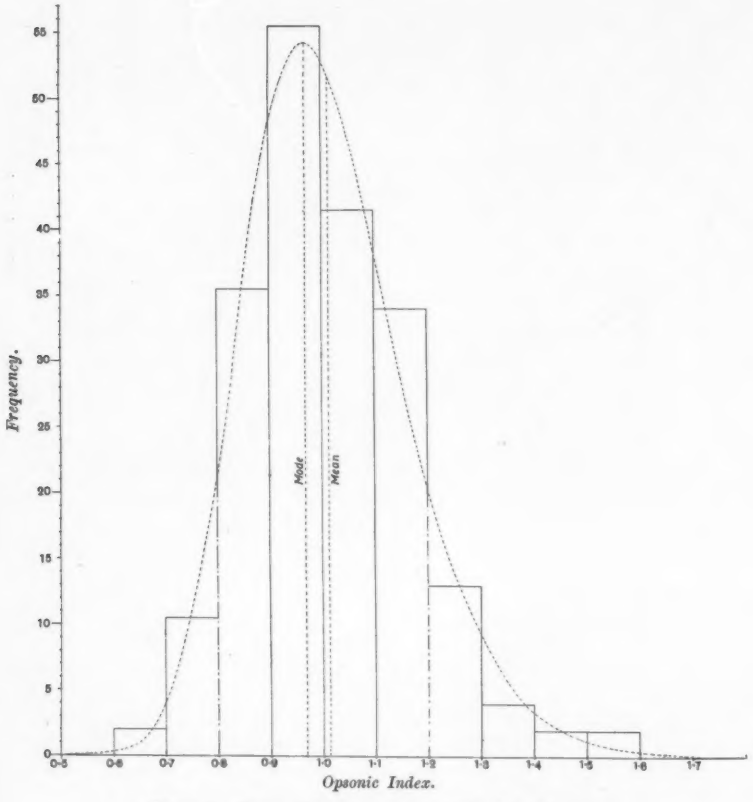
* See the following paper.



GRAPH 7. Curves of 50's with replacement—continuous: without replacement—dotted.

The actual number of bacteria in each sample of 25, 50 or 100 cells was copied out on a separate slip of cardboard, so that we had 800, 400 and 200 cards corresponding to the samples of 25, 50 and 100 cells. The cards used were small pieces of correspondence cards which, being smooth, could be shuffled very easily.

With these cards the following experiments were performed. Taking the case of the 400 cards relating to the fifties, the cards were thoroughly shuffled and one drawn out. The number on the card was written down, the card *not* being replaced. The remainder were once more shuffled, another drawing was made and the process was repeated until the whole 400 had been drawn. The cards were then once more shuffled and the process repeated, the second set of numbers being written down one by one under the first set of values. Each number of the first set was then divided by the corresponding number in the second set and the 400



GRAPH 8. Curve of 100's secondary without replacement.

indices thus obtained were analysed as usual. The same experiment was carried out in the case of 25's and 100's. Tables XI, XI A and Graphs 6, 7 and 8, contain the constants and data of the distributions and Table XII the chances of obtaining indices within assigned limits deduced from the areas.

TABLE XI A.

Experimental Determination of Index Distributions.

Index	Samples of 25 (Drawings not Replaced)		Samples of 50 (Drawings not Replaced)		Samples of 50 (Drawings Replaced)		Samples of 100 (Drawings not Replaced)	
	Observed	Calculated	Observed	Calculated	Observed	Calculated	Observed	Calculated
·3— ·4	1	·073 } 1·527	—	—	—	—		
·4— ·5	3	1·454 }	—	—	—	—		
·5— ·6	16·5	11·243	5	3·989	3	3·007	—	
·6— ·7	45	40·468	15·5	13·990	15·5	15·308	2	1·500
·7— ·8	83·5	89·850	33·5	37·969	34·5	39·863	10·5	11·697
·8— ·9	133·5	121·939	70	66·729	69	65·497	35·5	36·009
·9—1·0	130	134·204	79·5	80·213	71	77·683	55·5	52·780
1·0—1·1	105	122·028	66·5	71·759	78·5	72·001	41·5	46·583
1·1—1·2	83	96·635	52·5	51·973	59·5	54·817	34	29·003
1·2—1·3	55·5	69·188	37·5	32·676	28	35·425	13	14·177
1·3—1·4	57·5	46·012	23·5	18·792	22	19·873	4	5·850
1·4—1·5	37	28·990	6·5	10·249	13	9·830	2	2·145
1·5—1·6	19·5	17·560	5	5·452	3	4·334	2	1·00
1·6—1·7	11·5	10·339	1	2·870	1·5	1·717		
1·7—1·8	10·5	5·967	1	1·512	1·5	·814		
1·8—1·9	4	3·397	2	·803 } 1·735	—			
1·9—2·0	2	1·918	0	·432 }	—			
2·0—2·1	2	2·0	0	·5 }	—			
2·1—2·2	—	—	1		—			
2·2—2·3	—	—	—	—	—			
Totals	800	803·265	400	399·908	400	400·169	200	203·744

A comparison of these tables with nos. VIII—X indicates how far the limits of reliability are narrowed by a more complete study of actual conditions (see also Table XV *infra*).

It will be noticed that these experiments do not assume any special distribution for the original variates. They merely afford a partial answer to this question—if we had two sets of 400—200, 800—samples exactly the same as the 400 (200, 800) actually counted and formed at random 400 (200, 800) indices from them, one set always furnishing the denominator and no sample being used more than once, how are the resulting indices likely to be distributed? In strictness, therefore, these results are only applicable to our particular case. If, however, we assume that our counts are numerous enough and accurate enough to give the distributions of sample means of the size used, and for the emulsion used, with fair exactitude, then we can perform another experiment.

TABLE XII.

Chances of Obtaining Certain Deviations on the Basis of Three Experimental Index Distributions. (Method of Non-Replacement.)

Opsonic Index	Samples of 25		Samples of 50		Samples of 100	
	Fraction of Total Area bounded by corresponding Ordinate	Odds against such a Deviation or a Greater	Fraction of Total Area bounded by corresponding Ordinate	Odds against such a Deviation or a Greater	Fraction of Total Area bounded by corresponding Ordinate	Odds against such a Deviation or a Greater
·6	·0160	61·5 to 1	·0098	101 to 1	—	—
·7	·0665	14·0 to 1	·0448	21·3 to 1	·0075	132 to 1
·8	·1727	4·79 to 1	·1397	6·16 to 1	·0659	14·2 to 1
·9	·3247	2·08 to 1	·3067	2·26 to 1	·2460	3·07 to 1
1·1	·3546	1·82 to 1	·3135	2·19 to 1	·2571	2·89 to 1
1·2	·2336	3·28 to 1	·1835	4·45 to 1	·1121	7·92 to 1
1·3	·1475	5·78 to 1	·1019	8·81 to 1	·0413	23·2 to 1
1·4	·0899	10·1 to 1	·0549	17·2 to 1	—	—
1·5	·0537	17·6 to 1	·0293	33·1 to 1	—	—
1·6	·0317	30·6 to 1	·0157	62·7 to 1	—	—
1·7	·0188	52·2 to 1	—	—	—	—
Beyond the Limits *1·4—·6*	·1059	8·44 to 1	·0647	14·5 to 1	—	—
" " " *1·3—·7*	·2140	3·67 to 1	·1467	5·82 to 1	·0488	19·5 to 1
" " " *1·2—·8*	·4063	1·46 to 1	·3232	2·09 to 1	·1780	4·62 to 1
" " " *1·1—·9*	·6793	·47 to 1	·6202	·61 to 1	·5031	·99 to 1

Let us suppose that the distribution of samples of 50 that we obtained is actually the distribution of samples of 50 in general, then the following experiment may be considered. The 400 cards were shuffled and drawn as before except that after each drawing the card was replaced, the process being carried on until 400 numbers had been written down, a second series of 400 numbers was then obtained in the same way and the indices were computed. Tables XIII, XIV and Graph 7

TABLE XIII.

Constants of the Experimental Frequency Distribution Obtained by the Method of Replacement. (Mean, Mode and σ, Expressed in Terms of the Opsonic Index.)

Mean	Mode	σ	β_1	β_2	r	b	m_1	m_2
1·0196	·9697	·2056	·2077	3·2155	62·7551	44·6934	9·0844	51·6706

a_1	a_2	κ_2	β_2-3	Skewness	P
6·7337	37·9597	−·8536	·2155	·2428	·8988

Equation: $$y=80{\cdot}0808\left(1+\frac{x}{6{\cdot}7337}\right)^{9{\cdot}0844}\left(1-\frac{x}{37{\cdot}9597}\right)^{51{\cdot}6706}.$$

A change of ·1 in the index is taken as the statistical unit of grouping and index ·45 as zero.

TABLE XIV.

Chances of Obtaining Certain Deviations, on the Basis of an Experimental Curve from Samples of 50 (Method of Replacement).

Opsonic Index	Fraction of Total Area bounded by corresponding Ordinate	Odds against such a Deviation or a Greater
·6	·0075	132 to 1
·7	·0458	20·8 to 1
·8	·1454	5·88 to 1
·9	·3096	2·23 to 1
1·1	·3165	2·16 to 1
1·2	·1795	4·57 to 1
1·3	·0910	9·99 to 1
1·4	·0413	23·2 to 1
1·5	·0167	58·9 to 1
1·6	·0059	168 to 1

give the statistical constants and other information which resulted from an analysis of this experiment. It will be noticed that the table of chances does not differ appreciably from that due to the other experiment (XV). On this account, and because of the doubtfulness of the assumption in the case of 100's and the smaller practical interest in the case of 25's, the experiment was not repeated in these cases.

TABLE XV.

Comparison of Odds against Certain Index Deviations as Calculated on Different Assumptions for Samples of 50.

Index	Using True Mean	Experimental Curve without Replacement	Experimental Curve with Replacement
·7	88·3 to 1	21·3 to 1	20·8 to 1
·8	15·2 to 1	6·16 to 1	5·88 to 1
·9	3·38 to 1	2·26 to 1	2·23 to 1
1·1	3·63 to 1	2·19 to 1	2·16 to 1
1·2	12·0 to 1	4·45 to 1	4·57 to 1
1·3	39·3 to 1	8·81 to 1	9·99 to 1

Without in the least wishing to underestimate the shortcomings of both our material and method, it appears to us that Table XII can fairly be used as a rough practical test of the accuracy of the opsonic method in the case of the Tubercle Bacillus and that it confirms the opinions expressed in our former paper. A theoretical discussion of the index problem in general is reserved for another paper and more suitable material. It is now proper to consider very briefly some practical points.

In the first place, it has been asserted that the Tubercle Bacillus is an unfavourable organism to select for the purpose of studying the accuracy of the opsonic technic. While we cannot assert that no more favourable example could have been chosen, we are of opinion that the Tubercle Bacillus is certainly not the least satisfactory of the organisms commonly dealt with. For instance, Tubercle Bacilli act, or are acted on, almost entirely as single individuals, while Staphylococci or Streptococci occur in groups or chains respectively. These groups or chains are counted as so many individuals which, from the point of view of phagocytosis, may or may not be correct. The opsonic power required to attract a group of cocci to the leucocyte *may* not be greater than that associated with the ingestion of a single coccus. Under such conditions we have the difficulties as to the correct treatment of doublets and triplets—to which allusion has been made by Harvey and McKendrick—in an acute form.

Next as to the practical accuracy of the method for diagnostic purposes or in the control of dosage. The results here communicated, which are in agreement with our previous observations and also, *mutatis mutandis*, with those of Harvey and McKendrick, seem to demonstrate that single observations falling within the limits of 20 % on either side of the mean cannot, in the case of Tubercle Bacilli, be regarded as strongly suggestive of abnormality. If it is possible, as will often be the case when it is a question of diagnosis alone, to repeat the observation several times under comparable conditions, the opsonic method may be expected to yield results of a certain value. When, however, the opsonic index is used to

control injections of Tuberculin, so that no two determinations are made under the same conditions, we do not think the method has more than a very restricted field of usefulness. This leads us on to refer to a cognate matter.

Glynn and Cox use language* which seems to imply that we have attacked the opsonic index in the sense of denying that differences in phagocytic power as between normal and diseased persons occur. We have not asserted this nor does any statement of ours bear such an interpretation. We have merely pointed out, a contention not seriously challenged, that the claims advanced by certain enthusiasts to the effect that all variations of 10 % from the normal are probably or necessarily significant, must be entirely unjustifiable.

At the risk of being tedious, we desire to emphasise once more the distinction between vaccine therapy on the one hand and opsonic estimations on the other. If our work tends, as it certainly may tend, to unsettle the opinion of some readers as to the relative accuracy of the tuberculo-opsonic index, we wish in no way to be ranked among detractors of the claim of vaccine therapy to be regarded as a new and powerful weapon in the struggle against disease. On the contrary we are convinced of its present importance and look to the future for its still greater development.

In conclusion we would again remark that the results here published are in essential agreement with those communicated by Harvey and McKendrick and ourselves in recent numbers of *Biometrika*.

Conclusions.

1. Large phagocytic counts extending to 15,000 or 20,000 cells are markedly skew and exhibit signs of heterogeneity.

2. The exclusion of clumped bacilli has not been the main cause of the poor fit obtained.

3. The heterogeneity may be due to difficulties in counting.

4. The distributions of the means of samples of 25, 50 and 100 cells drawn from these populations are also markedly skew and excellently represented by Pearson's type curves.

5. This skewness cannot be due entirely to homotyposis.

6. The odds against a second sample of the same material giving an opsonic index, in terms of the true mean, outside of the limits 1·2 to ·8 are, in the case of samples of 25, about 3 to 1, in the case of 50's about 6 to 1 and for 100's about 15 to 1.

7. When the index is estimated on the basis of two small samples and without determining the mean of the normal series from a long count, the odds are reduced to 1·5 to 1, 2 to 1 and 4·6 to 1 respectively, for samples of 25, 50 or 100.

* Glynn and Cox: *Biochemical Journal*, Vol. IV. p. 300.

8. Even in the case of samples of 100, it is doubtful whether single determinations which give values even beyond the limits 1·3 to ·7 can be regarded as satisfactory evidence of differentiation.

9. The two previous conclusions are based on the analysis of four experimental frequencies.

10. The Tubercle Bacillus is not necessarily an unfavourable test of the opsonic method.

11. The opsonic index is less useful as a means of controlling treatment than for diagnostic purposes, since in the latter case it may be possible to make several comparable determinations.

12. Differentiation in respect of phagocytic power (Tubercle Bacilli) can possibly be established by the opsonic index in some cases.

ON THE CONSTANTS OF INDEX-DISTRIBUTIONS AS DEDUCED FROM THE LIKE CONSTANTS FOR THE COMPONENTS OF THE RATIO, WITH SPECIAL REFERENCE TO THE OPSONIC INDEX.

By KARL PEARSON, F.R.S.

(1) Given x and y two variables, the frequency constants of both of which are known, we often require in statistics the frequency constants of their ratio or index: $i = x/y$.

If the coefficients of variation are small, we have with the usual notation for means, standard deviations, moment coefficients, etc.*:

$$\bar{\imath} = \left(\frac{\bar{x}}{\bar{y}}\right)\left\{1 + v_x^2 - r_{xy}v_xv_y - \frac{\mu_3'}{\bar{y}^3} + \frac{p_{12}}{\bar{x}\bar{y}} + \frac{\mu_4'}{\bar{y}^4} - \frac{p_{13}}{\bar{x}\bar{y}^3} + \text{etc.}\right\} \quad \text{.....(i),}$$

$$M_2 = \sigma_2^2 = \left(\frac{\bar{x}}{\bar{y}}\right)^2\left\{v_x^2 + v_y^2 - 2v_xv_yr_{xy} + \frac{4p_{12}}{\bar{x}\bar{y}^2} - \frac{2p_{21}}{\bar{x}^2\bar{y}} - \frac{2\mu_3'}{\bar{y}^3} + \frac{3p_{22} - p_{11}^2}{\bar{x}^2\bar{y}^2} + \frac{3\mu_4' - 2\mu_2'^2}{\bar{y}^4} + \frac{2\mu_2'p_{11}}{\bar{x}\bar{y}^3} + \text{etc.}\right\} \text{.....(ii),}$$

$$M_3 = \left(\frac{\bar{x}}{\bar{y}}\right)^3\left\{\frac{\mu_3}{\bar{x}^3} - \frac{\mu_3'}{\bar{y}^3} - \frac{3p_{21}}{\bar{x}^2\bar{y}} + \frac{3p_{12}}{\bar{x}\bar{y}^2} + \frac{9p_{22} - 6p_{11}^2 - 3\mu_2\mu_2'}{\bar{x}^2\bar{y}^2} - \frac{3(p_{31} - \mu_2p_{11})}{\bar{x}^3\bar{y}} - \frac{9(p_{13} - \mu_2'p_{11})}{\bar{x}\bar{y}^3} + \frac{3(\mu_4' - \mu_2'^2)}{\bar{y}^4} + \text{etc.}\right\} \text{...(iii),}$$

$$M_4 = \left(\frac{\bar{x}}{\bar{y}}\right)^4\left\{\frac{\mu_4}{\bar{x}^4} + \frac{\mu_4'}{\bar{y}^4} + \frac{6p_{22}}{\bar{x}^2\bar{y}^2} - \frac{4p_{31}}{\bar{x}^3\bar{y}} - \frac{4p_{13}}{\bar{x}\bar{y}^3} + \text{etc.}\right\} \quad \text{.....(iv).}$$

These formulae go to the order of the 4th power in the coefficients of variation, but of course this is not to the same order of approximation in M_2, M_3 and M_4.

(2) It will be seen at once that these approximate formulae would be practically unworkable if x and y were correlated, as we should have to find 3rd and 4th order product moments.

* A rule denotes a mean value, σ a standard deviation, v, $=\sigma/\text{mean}$, is a coefficient of variation; μ_2, μ_3, μ_4 are the moment coefficients for x, μ_2', μ_3', μ_4' for y, M_2, M_3, M_4 for the index i, and $p_{uv} = S(x - \bar{x})^u(y - \bar{y})^v/N$, where N is the total number of pairs. Thus $p_{20} = \mu_2$, $p_{02} = \mu_2'$, etc.

If they be uncorrelated we have:

$$\bar{\imath} = \left(\frac{\bar{x}}{\bar{y}}\right)\left\{1 + v_x^2 - \frac{\mu_2'}{\bar{y}^2} + \frac{\mu_4'}{\bar{y}^4} - \text{etc.}\right\} \quad \text{.....(i)}^{\text{bis}},$$

$$M_2 = \sigma_2^2 = \left(\frac{\bar{x}}{\bar{y}}\right)^2\left\{v_x^2 + v_y^2 - \frac{2\mu_2'}{\bar{y}^2} + 3v_x^2 v_y^2 + \frac{3\mu_4' - 2\mu_2'^2}{\bar{y}^4} + \text{etc.}\right\} \quad \text{.....(ii)}^{\text{bis}},$$

$$M_3 = \left(\frac{\bar{x}}{\bar{y}}\right)^3\left\{\frac{\mu_3}{\bar{x}^3} - \frac{\mu_3'}{\bar{y}^3} + \frac{6\mu_2\mu_2'}{\bar{x}^2\bar{y}^2} + \frac{3(\mu_4' - \mu_2'^2)}{\bar{y}^4} + \text{etc.}\right\} \quad \text{.....(iii)}^{\text{bis}},$$

$$M_4 = \left(\frac{\bar{x}}{\bar{y}}\right)^4\left\{\frac{\mu_4}{\bar{x}^4} + \frac{\mu_4'}{\bar{y}^4} + \frac{6\mu_2\mu_2'}{\bar{x}^2\bar{y}^2} + \text{etc.}\right\} \quad \text{.....(iv)}^{\text{bis}}.$$

Formulae (i) to (iv) are extensions of formulae given by me in a paper on *spurious correlation**. Formulae (i)$^{\text{bis}}$ to (iv)$^{\text{bis}}$ are due to Dr M. Greenwood, Jun., who obtained them in dealing with the problem of the distribution of the opsonic index. They show at once two noteworthy but not yet noted points, namely, (*a*) if the distribution of both x and y be symmetrical, i.e. μ_3 and $\mu_3' = 0$, M_3 will not be zero or the distribution of indices must be skew; (*b*) the mean of the ratio of two numbers picked out of the same series is certainly greater than unity if the series be symmetrical, and will probably be always greater than unity even if it be not, i.e. $\bar{\imath} > \bar{x}/\bar{y}$, which is unity for the same x and y series†. Dr Greenwood found, however, that these formulae did not give with sufficient accuracy the constants of the index distribution. This was probably due to two causes: (*a*) clearly we ought only to keep to the square order in M_2 and the cubic order in M_3 if we retain only to the 4th order in M_4; or if we keep to the higher terms in M_2 and M_3, we must go further with M_4; and (*b*) the values of v_x or v_y are not so small, that the convergency is sufficient when we take these lowest terms of the expansions. It seemed accordingly desirable to find some other way of attacking the problem, and Dr Greenwood asked me for suggestions. The problem he had in view was the distribution of the opsonic index when the blood of the same individual taken in the same manner at the same time was treated as test and as normal. If a wide range of values of the opsonic index could thus be obtained, it would cast some light on what deviations from unity must be looked upon as significant, when test and normal were different individuals.

(3) The idea that suggested itself to me was a fairly simple one, namely to tabulate the y-frequencies to a variable $z = 1/y$. The units of the z-frequency groups will not be equal, but they will all be sufficiently small for us to concentrate their frequencies at their mid-points. We can then calculate their moments easily. Let $\nu_1, \nu_2, \nu_3, \nu_4$ be the moments of x about the zero value of x, and $\nu_1', \nu_2', \nu_3', \nu_4'$ be the moments of z about the zero value of z. Then

$$i = z \times x,$$

* *R. S. Proc.* Vol. LX. p. 492.

† Given two dice, it would be advantageous to bet that the ratio of the number of pips on the two at a cast will exceed unity.

and if m_1, m_2, m_3, m_4 be the moments of i about its zero value, we have:

$$m_1 = P_{11}, \quad m_2 = P_{22}, \quad m_3 = P_{33}, \quad m_4 = P_{44},$$

where P_{uv} is the uvth product moment of z and x about axes through their zero values. In the special case in which z and x are uncorrelated, as in the opsonic index,

$$m_1 = \nu_1\nu_1', \quad m_2 = \nu_2\nu_2', \quad m_3 = \nu_3\nu_3', \quad m_4 = \nu_4\nu_4',$$

or we can obtain any moment about the zero of i by multiplying the corresponding moments of x and z.

These moments are then transferred by the usual formulae

$$M_2 = m_2 - m_1^2, \quad M_3 = m_3 - 3m_2m_1 + 2m_1^3, \quad M_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$$

to the mean as origin and the type of frequency calculated in the usual way from the corresponding β_1 and β_2.

In Greenwood and White's data we have, for the three series discussed below, elementary subranges rising by ·32, ·24 and ·16 of a bacillus per leucocyte for the distribution of the means of counts of 25, 50 and 100. I find that for such distributions, the value of the variate z will only be affected by about a unit in the third place of decimals in the worst cases, i.e. the lowest values of y in samples of 25, whether we use for z (i) the mean of the inverses of the start and finish of the subrange, (ii) the mean of the inverses of all the 32, 24, or 16 hundredths in the subrange, or (iii) the inverse of the mid-point of the subrange. I have accordingly adopted the last as the simplest value of z for practical purposes.

(4) *Illustration of the method.* I. *Greenwood and White's* 200 *samples of* 100 *counts.*

The data are given in Table I. The moments of the frequency distributions for y and z as variates about the zero of those variates were then found by tables of powers of numbers and a calculating machine*.

I. *Distribution of* 40,000 *indices for* 200 *samples of* 100 *counts.*

For x: $\nu_1 = 3{\cdot}67620$, $\nu_2 = 13{\cdot}67643$, $\nu_3 = 51{\cdot}50333$, $\nu_4 = 196{\cdot}40357$.

For z: $\nu_1' = 0{\cdot}275165$, $\nu_2' = 0{\cdot}076603$, $\nu_3' = 0{\cdot}021577$, $\nu_4' = 0{\cdot}0061494$.

These give:

$$m_1 = 1{\cdot}01156, \quad m_2 = 1{\cdot}04766, \quad m_3 = 1{\cdot}11126, \quad m_4 = 1{\cdot}20776,$$

which transferred to the mean give for frequency constants of the 40,000 possible indices:

$\mu_2 = {\cdot}02440$,	$\sigma = {\cdot}1562$,	$\kappa = {\cdot}2651$,
$\mu_3 = {\cdot}00213$,	$\beta_1 = {\cdot}3123$,	Mean $= 1{\cdot}01156$,
$\mu_4 = {\cdot}00235$,	$\beta_2 = 3{\cdot}9472$,	Mode $= {\cdot}9774$.

The distribution is accordingly of Type IV.

* I have cordially to acknowledge help from Alice Lee, D.Sc., Julia Bell, M.A., and Amy Barrington, who have each worked out nearly the whole of one distribution for me, and from H. Gertrude Jones, who has prepared the diagrams.

TABLE I. *Greenwood and White's distribution of* 200 *samples of* 100 *counts. Opsonic Index. Tuberculosis.*

Variate x ...	*2·745*	*2·905*	*3·065*	*3·225*	*3·385*	*3·545*	*3·705*	*3·865*	*4·025*	*4·185*	*4·345*	*4·505*	*4·665*	*4·825*	*4·985*	*5·145*	Total
Frequency ...	2	9	4	13	29	36	34	34	15	11	5	3	1	1	2	1	200
Variate z ...	·364	·344	·326	·310	·295	·282	·270	·259	·248	·239	·230	·222	·214	·207	·201	·194	—

TABLE II. *Greenwood and White's Distribution of* 400 *samples of* 50 *counts. Opsonic Index. Tuberculosis*

Variate x ...	*2·35*	*2·59*	*2·83*	*3·07*	*3·31*	*3·55*	*3·79*	*4·03*	*4·27*	*4·51*	*4·75*	*4·99*	*5·23*	*5·47*	*5·71*	*5·95*	Total
Frequency ...	2	9	21	30	58	83	71	64	31	15	10	3	1	—	1	1	400
Variate z ...	·426	·386	·353	·326	·302	·282	·264	·248	·234	·222	·211	·200	·191	·183	·175	·168	—

TABLE III. *Greenwood and White's Distribution of* 800 *samples of* 25 *counts. Opsonic Index Tuberculosis.*

Variate x ...	*1·90*	*2·22*	*2·54*	*2·86*	*3·18*	*3·50*	*3·82*	*4·14*	*4·46*	*4·78*	*5·10*	*5·42*	*5·74*	*6·06*	*6·38*	*6·70*	*7·02*	Total
Frequency ...	1	14	38	74	128	145	147	113	74	37	19	7	2	—	—	—	1	800
Variate z ...	·526	·450	·394	·350	·314	·286	·262	·242	·224	·209	·196	·185	·174	·165	·157	·149	·142	—

We find: $r = 16{\cdot}5111$, $\nu = -9{\cdot}9138$ (since μ_3 is positive),
$a = {\cdot}52743$, $y_0 = 8638{\cdot}6$,
and the distribution of the 40,000 indices possible is given by:

$$Y = 8638{\cdot}6\left(1 + \frac{X^2}{{\cdot}27818}\right)^{-9{\cdot}2556} e^{9{\cdot}9138 \tan^{-1}\frac{X}{{\cdot}52743}} \quad \text{.....(i),}$$

where Y is the frequency at distance X from the origin of the curve which is at the opsonic index ·6949, the unit of X being absolute opsonic index measurement.

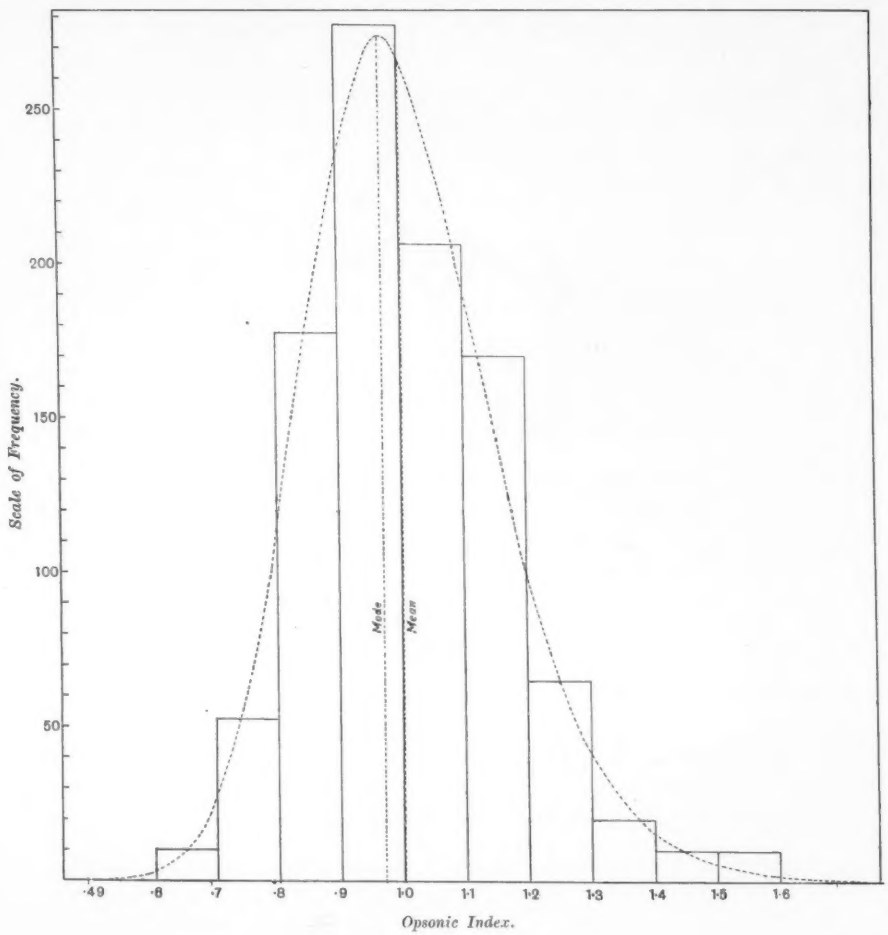


DIAGRAM I. Samples of 100 counted.

It will be seen, in accordance with the results drawn from the approximate formulae, that the mean value of the opsonic index is greater than unity and that there is marked skewness, i.e. no approach even with samples of 100 to a Gaussian distribution. Clearly the most probable value of the index is less than unity, and is rather farther from unity than the mean.

Greenwood and White have made an experimental determination by 200 random drawings. In Diagram I the frequencies of the above curve are reduced to 1000 total and plotted against their experimental data increased to 1000. This of course exaggerates the apparent deviations, but enables our three diagrams to be compared among themselves.

(5) *Further Illustrations.* The following frequencies are also deduced from Greenwood and White's results for 400 samples of 50 and 800 samples of 25.

With the same notation as before the constants of the distribution of the opsonic indices are as follows:

II. *Distribution of* 160,000 *indices for* 400 *samples of* 50 *counts.*

For x: $\nu_1 = 3{\cdot}68140$, $\nu_2 = 13{\cdot}82272$, $\nu_3 = 52{\cdot}92733$, $\nu_4 = 206{\cdot}69803$.

For z: $\nu_1' = {\cdot}277213$, $\nu_2' = {\cdot}078453$, $\nu_3' = {\cdot}022683$, $\nu_4' = {\cdot}006705$,

$m_1 = 1{\cdot}02053$, $m_2 = 1{\cdot}08443$, $m_3 = 1{\cdot}20055$, $m_4 = 1{\cdot}38591$,

$\mu_2 = {\cdot}042948$, $\sigma = {\cdot}2072$, $\kappa = {\cdot}5415$,

$\mu_3 = {\cdot}006191$, $\beta_1 = {\cdot}4837$, Mean $= 1{\cdot}0205$,

$\mu_4 = {\cdot}007567$, $\beta_2 = 4{\cdot}1022$, Mode $= {\cdot}9611$.

The distribution is of Type IV:

$$r = 20{\cdot}86018, \qquad \nu = -22{\cdot}67044,$$

$$a = {\cdot}62536, \qquad y_0 = 16{\cdot}9740,$$

and the equation to distribution of the 160,000 indices possible is given by:

$$Y = 16{\cdot}9740 \left(1 + \frac{X^2}{{\cdot}39107}\right)^{-11{\cdot}43009} e^{22{\cdot}67044 \tan^{-1} \frac{X}{{\cdot}62536}} \quad \text{.....(ii).}$$

The origin of the curve is at the opsonic index ·3409.

Diagram II gives this theoretical distribution reduced to 1000 cases and compared with Greenwood and White's experimental data for 400 increased to 1000 also.

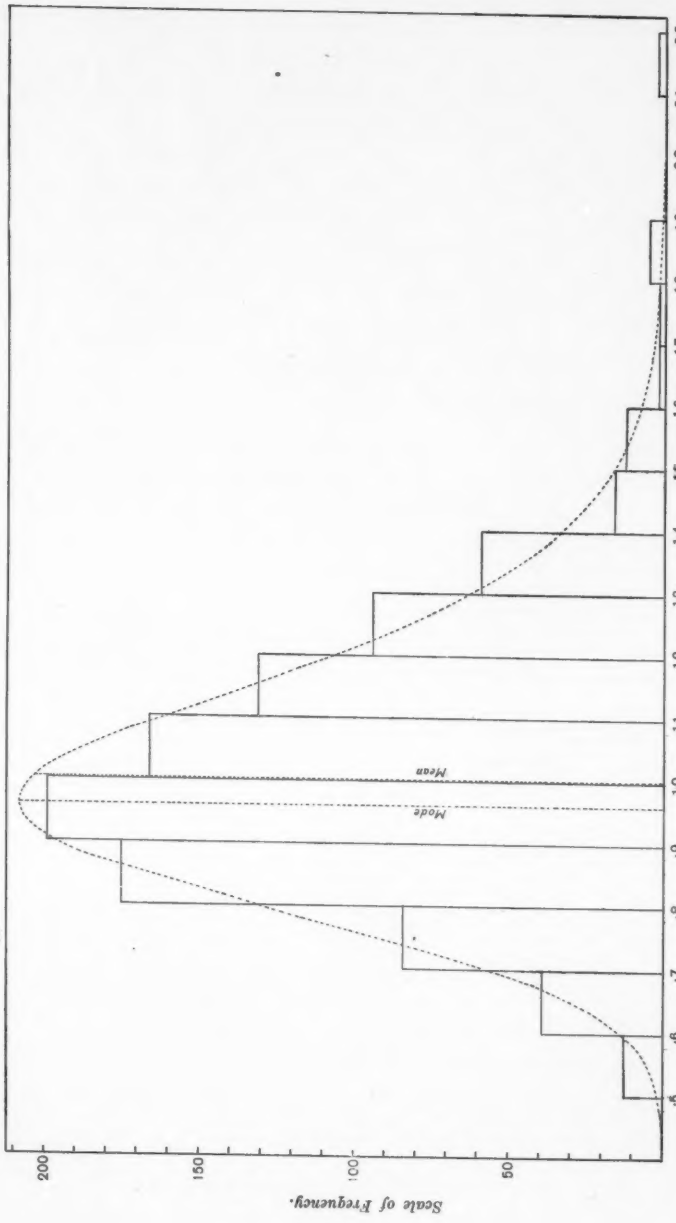


DIAGRAM II. Samples of 50 counted.

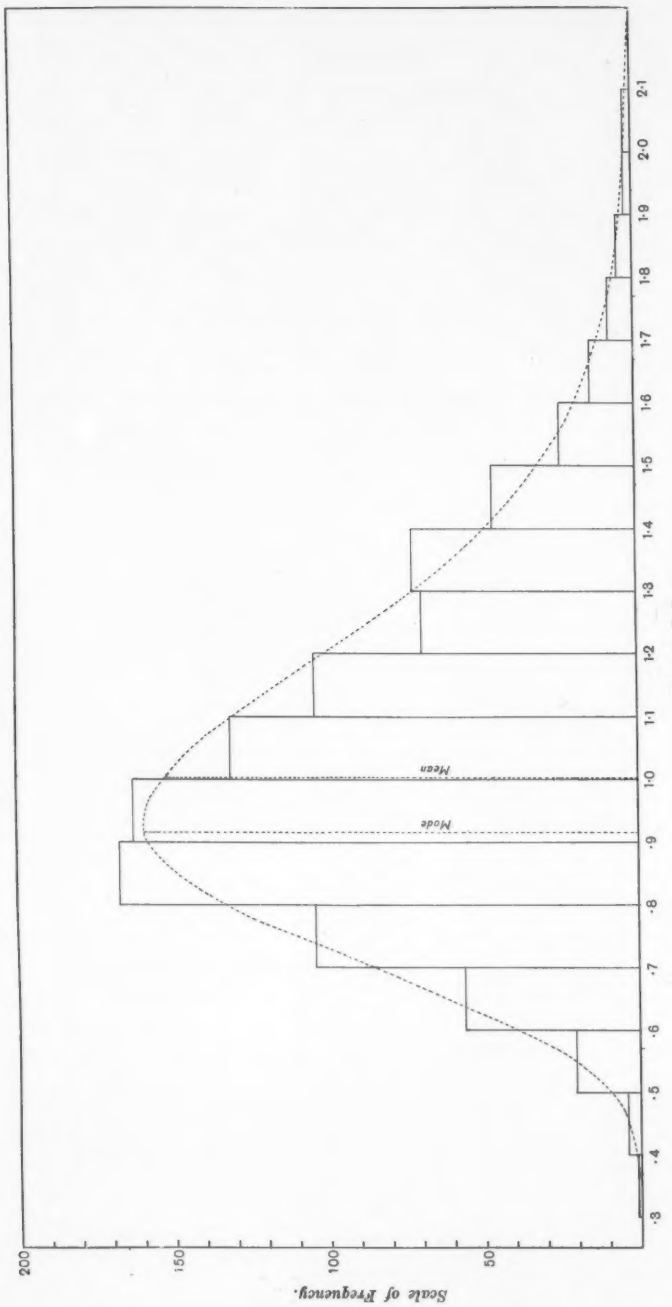


Diagram III. Samples of 25 counted.

III. *Distribution of* 640,000 *indices for* 800 *samples of* 25 *counts.*

For x: $\nu_1 = 3{\cdot}68160$, $\nu_2 = 14{\cdot}00990$, $\nu_3 = 55{\cdot}03056$, $\nu_4 = 222{\cdot}90230$,

For z: $\nu_1' = {\cdot}281298$, $\nu_2' = {\cdot}082055$, $\nu_3' = {\cdot}024861$, $\nu_4' = {\cdot}007834$,

$m_1 = 1{\cdot}03563$, $m_2 = 1{\cdot}14959$, $m_3 = 1{\cdot}36811$, $m_4 = 1{\cdot}74622$,

$\mu_2 = {\cdot}077063$, $\sigma = {\cdot}2776$, $\kappa = 1{\cdot}2480$,

$\mu_3 = {\cdot}017943$, $\beta_1 = {\cdot}7060$, Mean $= 1{\cdot}0356$,

$\mu_4 = {\cdot}025583$, $\beta_2 = 4{\cdot}3088$, Mode $= {\cdot}9330$.

The distribution is of Type VI*:

$$r = -31{\cdot}25271, \qquad q_1 = 51{\cdot}67822, \qquad q_2 = 18{\cdot}42551,$$

$$a = 1{\cdot}52075, \qquad \log y_0 = 26{\cdot}6813462,$$

and the equation to the distribution of the 640,000 indices possible is given by:

$$Y = \text{antilog } 26{\cdot}6813462 \times (X - 1{\cdot}52075)^{18{\cdot}42551}\, X^{-51{\cdot}67822} \quad \ldots\ldots\ldots\text{(iii)}.$$

The origin is at the opsonic index $-1{\cdot}4304$.

Diagram III gives this distribution reduced to a total of 1000 indices and set against Greenwood and White's experimental curve for 800 increased to 1000. This exaggerates the apparent deviations, but enables us to compare with the samples of 50 and 100. Considering the relative paucity of Greenwood and White's drawings, and the fact that we have fitted our curves to their non-replaced material, whereas a random sample would probably be better represented by the replaced material, the fits seem fairly close. The actual goodness of fit test was not applied, as at the time of writing the paper means of drawing the curves on a large scale and mechanically integrating them were not accessible.

(6) It is as well to look at a combined table of the frequency constants of the three distributions given above.

No. of Counts average based on	No. in Distribution of Means	Mean	Mode	σ	β_1	β_2
25	800	1·0356	·9330	·2776	·7060	4·3088
50	400	1·0205	·9611	·2072	·4837	4·1022
100	200	1·0116	·9774	·1562	·3123	3·9472

It will be seen that these values form a very consistent relatively smoothly altering system. But the approach to normality is very slow. Even with counts of 100 the distribution of the opsonic index is markedly skew and platykurtic, and it would not be safe to treat the distribution as a normal curve†. In all cases the

* Type VI is adjacent to Type IV and in this case the curve is almost on the boundary line—i.e. Type V: see Rhind's Diagram, *Biometrika*, Vol. VII. p. 389.

† The reduction of β_1 and β_2 to the Gaussian values 0 and 3 is, of course, not at the same rate as if we averaged 25, 50 and 100 opsonic indices. What we are doing here is to average the number of bacilli in 25, 50 and 100 cells on which the two factors of the index are based—a very different process.

modal or most probable value of the opsonic index of an individual tested by himself is less than unity and his average value greater than unity.

If we actually suppose the distributions normal and varying round the means with the standard deviations given above we have the following results:

Significance laid on an event which will not happen more frequently than	Number of Counts		
	25	50	100
Once in ten trials ...	·58—1·48	·68—1·36	·75—1·27
Once in eight trials ...	·61—1·46	·70—1·34	·77—1·25
Once in six trials ...	·65—1·42	·73—1·31	·80—1·23

This table is to be read in the following sense: If an opsonic index were based on a count of 50 cells, then once in eight trials an individual tested against himself would have an index lying outside the limits ·70 to 1·34. Or, again, with 100 counts once in six trials an individual tested against himself would have an index lying outside the limits ·80 to 1·23. It may be added that for most of the purposes of practical life or of exact science we should not consider an "improbability" which could happen once in ten trials as marking a significant differentiation. Much greater degrees of improbability would be required. In the case of medicine, however, much less certainty may be demanded of a judgment, and probably no weight would be given to an isolated opsonic determination in *diagnosis*. Still the matter gives ground for pause, the opsonic index of the same material tested by itself has a very wide range round unity.

The following table will show how closely Greenwood and White's experimental determinations agree with our theoretical evaluation of the constants:

Constant of Distribution	Samples of 25		Samples of 50		Samples of 100	
	G. + W.	P.	G. + W.	P.	G. + W.*	P.
Mean	1·0362	1·0356	1·0205	1·0205	1·0120	1·0116
Mode	·9432	·9330	·9536	·9611	·9689	·9774
Standard Deviation	·2585	·2776	·2204	·2072	·1553	·1562
β_1	·7099	·7060	·6952	·4837	·3654	·3132
β_2	4·3914	4·3088	4·8945	4·1022	3·7138	3·9472

The only substantial divergences are in Greenwood and White's values for β_1 and β_2 in the case of samples of 50. I have been through my results again and can find no error. Dr Greenwood has been through his figures and finds slight

* They have used the boundary curve Type V in the case of samples of 100.

slips (see footnote, p. 521) in the values of β_1 and β_2. These are not sufficient to account for the divergence of their results from mine, and from the trend of their samples of 25 and 100. I am inclined to think the divergence is due to the presence of the outlying index 2·16 in their chance drawings—a result my curve shows to be exceedingly improbable not once in 3000 drawings, and the like of which does not occur in their samples of 50 (drawings replaced) (see p. 525). It would clearly be possible from their experimental method of drawing samples of 800, 400 and 200 indices to closely approach the corresponding theoretical distributions of 640,000, 160,000 and 40,000 represented in our curves. Our results then are in every way confirmatory of theirs, but place on a rather more satisfactory theoretical instead of experimental footing their deduction of index distributions. The general conclusion seems to be that except in the case of an extremely low or extremely high value of the opsonic index, little if any weight whatever ought to be placed on a *single* determination of this index. Hence the method would not be valid when applied to cases in which, owing to the evolution of a morbid process or the action of some drug, very few observations can be made under the same conditions, i.e. it must be of doubtful application in treatment. Further the concentration obtained by basing the index on a count of 100, rather than one of 50 or even 25, while sensible is not very rapid. It would require very large numbers—much beyond every-day practice—to reduce in a marked manner this variation of the opsonic index from unity, when an individual even is tested against himself. Generally the diagrams indicate that an extreme variation in excess is more likely to occur than an extreme variation in defect, but that the most probable index, when the individual is tested against himself, will be one somewhat less than unity.

MISCELLANEA.

I. On the Value of the Teachers' Opinion of the General Intelligence of School Children.

Compiled from the Tables and Reductions of
H. GERTRUDE JONES, University College, London.

A great deal of random criticism has recently been made of school-teachers' estimates of general intelligence. We have been told that those estimates are absolutely without value, that any correlations obtained from them are idle and that the personal equation involved is so great that no use can legitimately be made of them. Good illustrations of this type of criticism will be found in the notices of Heron's memoir on *The Influence of Defective Physique and Unfavourable Home Environment on the Intelligence of School Children*, which Mr G. U. Yule has considered it desirable to publish in two separate journals*. As it does not appear that those who criticise the teacher's estimate of general intelligence in this way have made any experimental inquiry into the matter themselves, it may be of value to publish an investigation of a preliminary character made some time ago as an attempt to ascertain whether such classifications are really idle. More ample researches will shortly be published, but it may be said in advance here that they serve to confirm the present result.

The data are taken from schedules filled in by over 20 Aberdeen teachers†, who were not specially prepared for the task. 249 boys were taken from 4 different schools; their ages ran from 6 to 14 and they belonged to classes which were termed "Infants," I, II, III, IV, V, and VI. They were judged not only by different teachers, but in different schools, and their mental capacity was appreciated in four groups: (i) Excellent = E., (ii) Good = G., (iii) Moderate = M., (iv) Dull = D. This is a pure appreciation of general intelligence by a variety of teachers in a variety of schools. Next we have taken the individual boy's place in class as shewn by examination results and divided by the number of boys in the class. This may be adopted as a measure of the boy's examinational intelligence. If the teacher's estimate of general intelligence be of small value we should expect that when allowance is made for difference of class and age there would be little relation between examinational test and general intelligence.

The following characters were taken out and tabled, "Mental Capacity," Age, Position in Class (measured as above stated) and Class or Standard, and the following six tables obtained.

* *Journal of the Statistical Society*, Vol. LXXIII. p. 547, and *School Hygiene*, Vol. I. p. 473.

† For the Royal Commission on Physical Training (Scotland). We have gratefully to acknowledge the loan of these schedules by Professor Matthew Hay.

TABLE I.

Mental Capacity.

Age	E	G	M	D	Totals
6	4	11	11	5	31
7	2	15	10	6	33
8	5	19	6	4	34
9	3	18	12	—	33
10	2	20	11	1	34
11	2	9	12	5	28
12	1	11	10	3	25
13	—	9	13	2	24
14	—	4	3	—	7
Totals	19	116	88	26	249

TABLE II.

Mental Capacity.

Standards	E and G	M and D	Totals
Infants, I, II, III	77	61	138
IV, V, VI ...	58	53	111
Totals	135	114	249

TABLE III.

Class or Standard.

Age	Infants	I	II	III	IV	V	VI	Totals
6	30	1	—	—	—	—	—	31
7	—	30	2	1	—	—	—	33
8	—	12	21	1	—	—	—	34
9	—	3	3	21	6	—	—	33
10	—	—	—	10	22	1	1	34
11	—	—	1	—	4	22	1	28
12	—	—	—	2	1	9	13	25
13	—	—	—	—	—	7	17	24
14	—	—	—	—	—	—	7	7
Totals	30	46	27	35	33	39	39	249

TABLE IV.

Mental Capacity.

Place in Class	E	G	M	D	Totals
·03	7	9	2	—	18
·08	3	8	—	—	11
·13	4	10	1	—	15
·18	3	12	2	—	17
·23	1	12	3	1	17
·28	—	11	1	—	12
·33	—	10	3	—	13
·38	—	7	4	—	11
·43	—	5	5	—	10
·48	—	11	9	—	20
·53	—	8	5	1	14
·58	1	5	8	1	15
·63	—	2	7	—	9
·68	—	4	6	—	10
·73	—	2	4	3	9
·78	—	—	8	1	9
·83	—	—	11	—	11
·88	—	—	5	5	10
·93	—	—	2	6	8
·98	—	—	2	8	10
Totals	19	116	88	26	249

TABLE V.

Age.

Place in Class	6	7	8	9	10	11	12	13	14	Totals
·03	2	3	—	5	3	2	3	—	—	18
·08	3	1	—	2	2	1	1	1	—	11
·13	3	3	1	3	1	—	1	2	1	15
·18	4	1	2	3	3	1	—	3	—	17
·23	1	3	6	1	1	2	—	3	—	17
·28	—	2	3	1	4	—	1	1	—	12
·33	—	—	4	3	4	—	2	—	—	13
·38	2	1	—	—	3	1	2	—	2	11
·43	—	2	1	1	1	2	2	1	—	10
·48	3	1	3	1	1	4	4	2	1	20
·53	1	3	3	3	—	3	—	—	1	14
·58	1	—	3	2	1	4	2	2	—	15
·63	1	2	1	1	4	—	—	—	—	9
·68	—	1	2	2	3	—	—	2	—	10
·73	1	3	2	—	—	2	1	—	—	9
·78	2	2	1	2	1	—	—	1	—	9
·83	1	1	—	2	1	2	1	2	1	11
·88	—	1	1	1	1	2	1	2	1	10
·93	2	1	1	—	—	1	2	1	—	8
·98	4	2	—	—	—	1	2	1	—	10
Totals	31	33	34	33	34	28	25	24	7	249

TABLE VI.

Class or Standard.

Place in Class	Infants	I	II	III	IV	V	VI	Totals
·03	2	3	—	5	5	2	1	18
·08	3	1	—	1	2	2	2	11
·13	3	4	1	2	1	2	2	15
·18	4	2	2	2	2	2	3	17
·23	1	3	4	3	—	3	3	17
·28	—	3	2	1	5	—	1	12
·33	—	1	4	5	1	1	1	13
·38	2	—	2	1	2	1	3	11
·43	—	2	1	1	1	3	2	10
·48	3	2	2	1	1	5	6	20
·53	1	6	—	3	1	2	1	14
·58	1	1	2	1	3	5	2	15
·63	1	1	2	2	3	—	—	9
·68	—	2	1	2	3	1	1	10
·73	1	5	—	—	1	2	—	9
·78	1	3	1	2	1	—	1	9
·83	1	3	—	1	—	2	4	11
·88	—	1	2	1	1	3	2	10
·93	2	1	1	1	—	1	2	8
·98	4	2	—	—	—	2	2	10
Totals	30	46	27	35	33	39	39	249

Let us look at the results of these Tables in succession:

Table I. Mental Capacity and Age. The mean ages are as follows:

E: 8·87
G: 9·89
M: 10·33
D: 9·38
} All classes: 9·91.

There is thus a decrease of intelligence with age until we come to the worst class and here there is a drop in age. There are very certainly cross-currents at work, e.g. clever boys go young to school, and leave early, and extreme cases of dullness go to "special" schools at 9 to 11, and school lessens at any rate the outward appearance of stupidity. Worked out by the correlation ratio method the correlation is $\eta = \cdot 1830$. This clearly should be considered as negative, but while the regression is not linear and therefore the correlation ratio sensibly greater than the correlation coefficient, we will take $r = -\cdot 1830$ because we believe the "dullness" of the younger children to some extent wears off with continued school life or the dull are drafted elsewhere*.

* The reader should examine the last column of Table I. We think several of the younger children would ultimately pass into "special schools" or be sharpened by school life. The same bi-modal character is obvious in column M, if to a lesser extent.

Table II. Mental Capacity and Standard. It is interesting to look at the data for this as a contingency table. This is given below as Table II *bis* :

TABLE II *bis.*

Class or Standard.

Mental Capacity.	Infants	I	II	III	IV	V	VI	Totals
E	4	5	3	2	1	3	1	19
G	11	19	15	18	21	10	22	116
M	10	14	7	15	10	18	14	88
D	5	8	2	—	1	8	2	26
Totals	30	46	27	35	33	39	39	249

The greatest contributions to the contingency arise from (i) the defect of Dulls in Standard III and (ii) the excess of Dulls of Standard V, accompanied as it is by a defect of Good mental capacities. These may be personal to the teachers of these standards, or to their selection of children, which was made for the purposes of physical measurement. But the corrected contingency coefficient which should be much emphasised by personal equation is only ·2123*. Now we shall see that the correlation between age and standard is very high, i.e. ·9366; hence the correlation between mental capacity and age for constant standard $= -·0461$, or is practically negligible. In other words, allowing for the personal equation of the master with regard to intelligence, age counts for very little in his estimate of general intelligence. If we even out the irregularities of the above table by forming out of it a mere four-fold table as in Table II we obtain $r = -·0558$. We expect that this is really the more correct measure of the true relationship between standard and general intelligence.

Table III. Age and Standard. The correlation here is clearly very high ; determined as a correlation ratio it is ·9366. Practically children are moved up a class or standard a year.

Table IV. Mental Capacity and Place in Class. The correlation as found by correlation ratio is ·7246. The actual mean places of each category are :

E : ·122
G : ·312
M : ·500
D : ·851

} All categories : ·456.

This list itself shews that the teacher's estimate of general intelligence represents very closely the examinational place which a student will take and for many inquiries may quite legitimately replace that estimate of ability.

Table V. Age and Place in Class. Found by the ordinary product moment method the correlation coefficient is ·0484, shewing that with increased age a *slightly worse* class position is taken, i.e. any advantage from age is counterbalanced by the clever elder boys leaving school earlier. The value, considering the numbers dealt with, is, however, hardly significant.

Table VI. Place in Class and Standard. The mean places are as follows :

Infants ·457
I ·494
II ·445
III ·393
IV ·388
V ·493
VI ·493

} All classes : ·456.

* This must be taken negative if treated as a correlation coefficient.

This series is very irregular. The η for the series is ·1553, but this is rather a measure of the irregularity than of anything else*. If we take the correlation coefficient for a four-fold table as follows

	Infants I, II, III	IV, V, VI	Totals
·05 to *·45*	71	53	124
·50 to *1·00*	67	58	125
Totals	138	111	249

we find $r = ·0580$, a number of scarcely any practical significance and shewing that really place in class is, as it should be, independent of standard.

We have then to sum up the following results:

Intelligence and Age . . .	$r_{13} = -·1830$
Intelligence and Position in Class .	$r_{12} = ·7246$
Intelligence and Standard . .	$r_{14} = -·0558\,(-·2123)$
Position in Class and Age . .	$r_{23} = -·0484$
Position in Class and Standard .	$r_{24} = ·0580\,(·1553)$
Age and Standard	$r_{34} = ·9368$

In the case of two of these correlations for which some doubt might exist, we have placed in brackets the higher values found by different processes. The results for both values of the partial correlation coefficient have been worked out.

Correlation of "General Intelligence" and Examinational Place, for constant age and constant standard:

With lower values of r_{14} and r_{24}: $\rho_{12} = ·70 \pm ·02$.
With higher values of r_{14} and r_{24}: $\rho_{12} = ·99 \pm ·01$.

We may safely draw from these results the conclusion that the teacher's estimate of general intelligence is not a purely idle character, wholly valueless owing to the personal equation of the teacher. Whatever else it may be, it is highly correlated with the place which the student will take in an examination test and accordingly has at least one quite definite significance.

The first of these results shews that with the lesser and more probable values of the correlations we obtain, allowing for standard and age, a high correlation (·70) between the teacher's estimate of general intelligence and actual examination measure of capacity. There are, however, cross-currents at work in elementary schools; the one is the selection of the notably dull children about 10 who are drafted into "special schools." This causes rather a defect of dull children in the II, III and IV standards; the next is the selection of the more intelligent children in the highest standards (this is most obvious in V) to leave the elementary schools. The total effect is to make a somewhat high correlation ratio for standard and intelligence. If we took all the children in a class there ought to be sensibly no correlation between position in class and standard, but we actually find for classes II, III and IV excess of good places (averages ·445, ·393, ·388 instead of ·456) and for the high classes V and VI excess of bad places (averages ·493, ·493). This seems to be the result of the same cross-currents, the selection of the

* It must be remembered that not *every* child, but only 30 to 40 in each standard were physically examined, and the number so dealt with is not by any means the number in the class. This selection of children accounts for the mean place in class being ·456 and not ·500.

dull to go and of the dull to stay at different ages. If we now give the higher values to the correlations produced by these tendencies in elementary schools, we find on correction not that they have tended to disguise the real value of the "general intelligence" character by giving it a spurious high intensity, but that they have tended to *lower* its true importance. Those who complain of the different percentages of the divers grades of intelligence of the children in the different standards of the same elementary school often seem unaware that the same sample does not pass through all standards, at one stage there is a rejection of the very dull and at another the retention of the dull class. The main facts brought out are the substantial correlation between teacher's estimate of general capacity and examination test, and the fact that extreme allowance for age and standard tends to emphasise this relationship rather than to shew that the teacher's estimate is of little value. These observations were not made *ad hoc*, but they have been confirmed by other observations made on far larger numbers with additional safeguards for accuracy. The results of these will shortly be published, but it seems worth while indicating at a time when hasty criticisms are being made of the value of "general intelligence" estimates that the teacher's appreciation of mental capacity does mean something, and has a very direct and practical value, when properly registered and handled.

K. P.

II. Note on the Separate Inheritance of Quantity and Quality in Cows' Milk.

By KARL PEARSON, F.R.S.

Under the above title Professor James Wilson, of the Royal College of Science, Dublin, read a paper in May, 1910, which is published in Vol. XII. pp. 470—479, of the *Scientific Proceedings of the Royal Dublin Society.*

He states that: "It is a very general opinion that the milk of high-yielding cows is usually poorer and that of low-yielding cows richer in quality" (p. 470), and again he writes:

"If we group together all the low-yielding cows, and find their milk *invariably* high in quality, we may infer that low yield and high quality are of the nature of concomitant variations. If we group the high-yielding cows together, and find their milk *invariably* of low quality, we may infer that high yield and low quality run together. But if we take these groups and any other groups we can form, and find that the quality *varies the same way in them all*—that is that there are low qualities, high qualities, and medium qualities in every one of them—then we are *justified in inferring* that the quantity and quality of the milk are independent of each other. And this is what we do find" (p. 471).

The italics are mine. It would be difficult to find a paragraph containing more fallacies in as many words. Two qualities may be associated together, even closely correlated, and yet one will not be *invariably* accompanied by the other. High, low and medium values of one character may occur with any given value of a second and yet there be high correlation between them. The whole problem turns on the *extent* of the variation of the first character for a given value of the second, and not the invariable appearance of one with the other. The author may have some conception of correlation other than absolute association, but neither this paragraph, nor his general treatment of the subject, shews any signs of it.

Professor Wilson's material is involved in the table below.

Now there can be no doubt of the value of such returns and of tabling them, although we regret the clubbing together of the tail frequencies, that common error of the non-statistical trained mind, which renders determination of the true correlation so difficult.

Percentages of Butter-Fat in Milk of Ayrshire Cattle.*

Yield of milk	2·6	2·7	2·8	2·9	3·0	3·1	3·2	3·3	3·4	3·5	3·6	3·7	3·8	3·9	4·0	4·1	4·2	4·3	4·4	4·5	4·6	4·7	4·8	4·9	5·0	5·1	Totals
Under 500 gallons	1	—	4	10	7	16	17	26	56	66	52	67	69	61	36	30	23	16	9	6	7	3	1	—	1	1	585
From 500 to 600 gallons	—	—	4	7	6	21	41	63	99	103	107	125	105	89	67	51	25	18	4	3	7	1	1	1	—	—	948
From 600 to 700 gallons	1	—	3	4	13	17	35	63	64	103	106	99	86	62	54	45	27	10	5	2	7	2	—	1	—	—	809
Over 700 gallons	—	—	—	1	8	14	31	38	52	59	66	70	61	43	35	19	15	9	2	—	1	—	—	—	—	—	524†
Totals ...	2	—	11	22	34	68	124	190	271	331	331	361	321	255	192	145	90	53	20	11	22	6	2	2	1	1	2866

* Data taken from *Report of Milk Records for Season* 1908, Ayrshire Cattle Milk Records Committee.

† Professor Wilson has 525, but only 524 cows appear in his printed table.

How now does Professor Wilson extract from these data the independence of quality and quantity of milk on which he insists, and which is chosen as the very title of his paper? He has prepared six elaborate diagrams, and he leaves the reader from mere inspection of these to come to the same conclusion as himself, i.e. that the variation is the same in all his groups and the quality independent of the quantity. Not being able myself to judge from such diagrams whether there is complete independence between quality and quantity, I thought it worth while to ascertain the mean percentages of fat and the variations in fat of his four groups. The results are as follows, the probable errors being added:

Quantity of Milk	Means	Standard Deviations
Under 500 gallons ...	3·710 ± ·010	·372 ± ·007
From 500 to 600 gallons ...	3·673 ± ·007	·319 ± ·005
From 600 to 700 gallons ...	3·651 ± ·008	·321 ± ·005
Over 700 gallons	3·640 ± ·008	·275 ± ·006

The means and variabilities are in percentage of butter-fat.

For the whole population of 2866 cows the mean is 3·672 and the variability is measured by 3·230. It will accordingly be clear that there is (i) a slight but perfectly definite decrease in quality with increase of quantity, the low-yielding cows have a higher, the high-yielding cows a lesser fat percentage than the average Ayrshire, although the increase and decrease only amount to about 1 °/₀ of the average yield, and (ii) the variation in quality is not the same for cows of each quantity category. Cows which give high quantity have far less variation in their yield of butter-fat than cows of low yield. The number of cows is so large relatively that the probable errors are small and there is no doubt about the results being significant. Owing to the manner in which the tails of the above frequency have been clubbed together it is not possible to work out the proper correlation by the ordinary product moment method. The means, however, suggest that the correlation cannot be very far from linear, and the correlation ratio η may be fairly taken to measure r. This gives us $r = -·075$ with a probable error of $\pm ·013$, a result, if small, yet definitely significant.

It seemed worth while putting these numerical results on record, as a further protest against somewhat dogmatic inferences drawn from mere inspection of elaborate but unnecessary graphs.

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS